Supporting Information

Supplement to: Sun S, Zhang W, Chen D, *et al*. **A glioma classification scheme based on co-expression modules of EGFR and PDGFRA**

Materials and Methods

Clustering analysis of glioma transcriptome data

Transcriptome data of gliomas (GSE4290, GSE16011, the REMBRANDT and the TCGA mRNA-seq data sets) and mouse brain development (GSE9566) were gathered from published studies (1-4), or from the TCGA data portal. The CEL files for the REMBRANDT data set (3) were merged and normalized with the RMAExpress tool. Except for Pearson correlation coefficient analysis, the expression values were expressed in the log2 scale in all analyses. The Tiantan data set was generated with 212 glioma samples collected at the Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University in Beijing from patients treated between 2006 and 2009. The use of glioma samples was approved by the Ethics Committee of Beijing Tiantan Hospital, and written informed consent was obtained from all patients. The gene expression data and the clinical and pathological parameters of all gliomas in the Tiantan data set can be retrieved from the Chinese Glioma Genome Atlas, www.cgga.org.cn, under data portal, EMPM online. All gliomas were diagnosed according to the 2007 WHO classification (5). Low-grade astrocytomas and oligodendrogliomas were treated with radiotherapy, GBMs with radiation therapy and alkylating agent-based chemotherapy. The longest follow-up was 56 months. The samples taken at surgery, were snap-frozen immediately in liquid nitrogen.

An haematoxylin and eosin stained frozen section was prepared for assessment of percentage of tumor cells. Samples with greater than 70% tumor cells were identified and selected. Total RNA from these frozen tumor tissue samples were extracted using the mirVana miRNA Isolation kit (Ambion, Austin, TX) according to the manufacturer's protocol. RNA concentration and quality were measured using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Houston, TX). The expression profiling of all samples was performed with Agilent Whole Human Genome Oligo Microarray in a 4*44k slide format. Each block represents more than 41,000 probe sets. All steps including sample-labeling, hybridization, washing and scanning were conducted at the laboratory in the Cancer Institute/Hospital of Peking Union Medical College, following the manufacturer's specifications. In brief, Cy3-labeled cRNA was generated from 500 ng total RNA using Agilent's Low RNA Input Linear Amplification Kit PLUS, and 1.65 μg cRNA from each labeling reaction was hybridized to one block of the microarray. After hybridization, slides were washed and then scanned with the Agilent G2565BA Microarray Scanner System. Fluorescence intensities on scanned images were extracted and preprocessed by Agilent Feature Extraction Software (v9.1). Data normalization and filtering was performed using GeneSpring GX 11.0 (Agilent Technologies). Only those genes with expression levels marked as "present" or "marginal" in all of the chips (blocks) passed through the quality filtering. Subsequently, measurements were set to less than 0.01, and the chip data was normalized to the 50th percentile of the intensity of each chip.

Qlucore Omics Explorer 2.2 (Qlucore AB, Lund, Sweden) was used for gene co-expression module construction and differential gene expression between the glioma

subtypes. Using Pearson correlation coefficient analysis in the GSE4290 data set, genes with their expression patterns highly correlated to EGFR and PDGFRA were identified as EM and PM, respectively. Unsupervised hierarchical clustering analysis in GSE4290 after variance filtering, which excluded all genes with standard deviations below 2.4% of the maximal standard deviation, was used to identify the 29 EM and 40 PM genes as the EM/PM classifier genes.

To define glioma clusters, we used non-negative matrix factorization (NMF) to analyze the expression data of the EM/PM classifier genes in each data set (6). The number of clusters in each data set was separately determined according to the following conditions: 1) the clusters were stable (the stability was measured by the cophenetic correlation coefficient between the consensus cluster matrix obtained by NMF and the hierarchical clustering based on the same); 2) these clusters shall be larger than 2. Pilocytic astrocytomas in GSE16011 were excluded in this analysis. *DMRTA2* and *LOC100507187* were not found in the Tiantan data set. Eight genes (*C1orf106*, *LIX1L*, *LOC100507187*, *LOC283174*, *OLIG1*, *PMS2P11*, *RNF180*, and *UHRF1*) were not detected in GSE16011.

To compare the EM/PM classification scheme with that of Phillips *et al*. (7) or Verhaak *et al*. (8), NMF analysis was performed in Tiantan, GSE16011 and the REMBRANDT data sets using the classifiers of Phillips *et al*. (7) or Verhaak *et al* (8). Samples in the resulting clusters were analyzed for survival and cross-labeled with EM/PM clusters in NMF analysis. Using supervised hierarchical clustering, comparison between EM/PM

classification scheme and the classification scheme of G-CIMP phenotype (9, 10) was performed with Qlucore explorer.

Analyses of expression signatures of genes characteristics of SVZ astrocytes, glial genesis and mature neural cell types in EM/PM gliomas

All these analyses were performed under the supervision of EM- and PM-defined clustering. Using Qlucore explorer, the expression signature of the M13C module of SVZ astrocytes (11) was examined in the Tiantan, REMBRANDT and GSE16011 data sets. The top 602 most significantly differentially expressed genes between mouse immature astrocytes (P1 to P8 astrocytes) and immature oligodendrocytes (non-myelinated oligodendrocytes and OPCs (4)) were identified using a *t*-test with significance threshold $p = 5.4 \cdot 10^{-6}$ and exported as developing astrocyte (371 genes) or developing oligodendrocyte classifier (231 genes). Using GSEA (12), these two classifiers (gene lists will be provided upon request.) were projected into the transcriptome data of EM/PM gliomas in the Tiantan, GSE16011 and the REMBRANDT data set to test for enrichment among the genes most highly differentially expressed between EM and PM subtypes. Similarly, the transcriptome data of the mature astrocytes, oligodendrocytes and neurons in GSE9566 were analyzed with unsupervised hierarchical clustering after variance filtering, excluding genes with standard deviations below 48% of the maximal standard deviation. Highly variably expressed genes for each mature cell type were exported as classifiers (gene lists will be provided upon request.). The enrichment of the signatures of the mature

astrocytes, oligodendrocytes and neurons in EM/PM glioma subtypes was analyzed in the Tiantan and the REMBRANDT data sets.

Analyses of SNP-array, gene dosage effect and gene modulation capacity

Copy number analysis was first performed on 205 gliomas from the REMBRANDT data set (3), which was based on 100 K SNP array. The raw HindIII CEL files for tumor and matched normal samples were imported into dChip (13). Copy number profiles were calculated by (a) normalizing the intensities using an invariant set of probes, (b) calculating a raw copy number for each SNP using "model based expression" with perfect match/mismatch difference, and (c) using median smoothing with a 10 SNP window. To eliminate batch specific noise in the copy number data the calculations were performed in smaller batches based on the creation date of the CEL-files. The copy number data was segmented and analyzed using GLAD (14) and GISTIC2.0 (15) at an amplitude threshold of ± 0.2 as implemented in the web application of Genepattern at http://genepattern.broadinstitute.org.

The copy number analysis based on SNP6.0 data was replicated in the TCGA mRNA-seq cohort on 349 gliomas. Normalized segmented copy number data were analyzed with GISTIC2.0 under the same amplitude threshold as above.

The gene dosage effect on the expression of EM/PM classifiers was analyzed in the TCGA mRNA-seq cohort in 123 EM and 133 PM gliomas. Spearman´s rank correlation analysis on the data of gene expression and gene dosage was performed. Odds ratio and Fisher´s exact test were used to analyze the correlation among the SCNAs of EM/PM classifiers

with gene dosage-dependent expression. To analyze the capacity of those EM/PM classifiers with gene dosage-dependent expression to affect the entire EM/PM signatures, linear regression analysis between the expression of EM/PM classifiers with gene dosage-dependent expression and EM/PM scores was first performed in the TCGA mRNA-seq cohort, and subsequently replicated in the REMBRANDT and GSE16011 data sets.

Survival analysis

Overall survival time was calculated from the date of surgery until death or the last follow-up contact. Kaplan-Meier survival curves were generated and analyzed with the log-rank test using Prism 5.0 software. Cox regression analysis was performed with SPSS software, with survival time as the dependent variable. EM and PM scores for each glioma sample were calculated as the average of normalized expression levels for EM or PM classifier genes detected in each data set, and used as variables in multivariate Cox regression analysis.

Legends for Supplementary Figures

Fig. S1. Identification of EM- and PM-based classifier genes and glioma molecular subtypes.

(**A**) Identification of EM/PM classifier genes. The top 37 and 44 known genes in EM and PM, respectively, were identified in GSE4290 by Pearson correlation coefficient analysis. After variance filtering, which excluded genes with standard deviations below 2.4% of the maximal standard deviation, 29 of the EM and 40 of the PM genes remained. These 69 genes were used as classifier genes throughout the study. Genes previously known to be relevant in glioma genesis are labeled in blue.

(**B**) Identification of glioma molecular subtypes. Gliomas were clustered using NMF of the EM/PM expression signatures in each data set. The consensus clustering matrix and cophenetic coefficients for the clusters are shown. K = 4, 5 and 4 were selected in Tiantan, GSE16011 and the REMBRANDT data sets, respectively. Under these conditions, all non-tumor samples were assigned to the same clusters.

Fig. S2. Validation of EM/PM glioma subtypes in the TCGA mRNA-seq data set.

Gliomas in the TCGA mRNA-seq data set were clustered using NMF of the EM/PM expression signatures at k = 5. Heat map of EM/PM signatures in glioma samples is shown in upper left. The plot of cophenetic coefficient analyses (upper right) shows the stability of the clusters. The composition of morphologically diagnosed gliomas in each EM/PM glioma subtype, the age at diagnosis, the survival profiles of patients, and the frequencies

of *IDH1* mutation, 1p36/19q13 co-deletion, co-alteration of *EGFR* and *PTEN*, focal amplification of *EGFR* and *PDGFRA*, and co-amplification of *EGFR* and *PDGFRA*, are summarized in lower left. Kaplan-Meier plot of the overall survival for the patients from each EM/PM subtype is shown in middle right. The overall survival data were analyzed using log-rank tests. The same color codes were used in the heat map and the Kaplan-Meier plot. Plot of principle component analysis of the whole transcriptome excluding genes with the standard deviation of their expression levels below 2.4% of the maximal standard deviation is shown lower right. The PM^{++}(Chr 7 Amp) gliomas showed similar genomic and transcriptomic profiles to EM gliomas.

Fig. S3. Definition of subclusters within EM and PM glioma subtypes.

The EM and PM scores for each glioma sample were calculated. The EM and PM gliomas were further sub-clustered based on the difference between the EM or PM scores.

Fig. S4. Highly similar subtype-specific transcriptome profiles for gliomas within each of EM, PM and $EM^{low}PM^{low}$ glioma subtypes.

(**A**) The Tiantan (upper), GSE16011 (middle) and REMBRANDT (lower) data sets were analyzed for whole transcriptome profile under the supervision of the EM/PM classification. Heatmaps at the indicated false discovery rate (FDR), identifying the 800 most variably expressed genes, are shown. Genes concordantly and differentially enriched in EM and PM glioma subtypes in these data sets are summarized in Table S4.

The numbers of EM/PM gliomas in each of the sample clusters are indicated. In all three data sets, the vast majority of gliomas clustered along the subtypes EM^{+++} and EM^{++}, or PM, or $EM^{low}PM^{low}$, respectively. In GSE16011, similarity between PM gliomas and normal brain tissues was evident. In Tiantan and REMBRANDT data sets, a subset of PM gliomas showed similarity to $EM^{low}PM^{low}$ gliomas. Another subset of PM gliomas were also enriched in genes that distinguished EM gliomas from other subtypes. A cluster in the REMBRANDT data set was found containing all EM/PM glioma subtypes.

(**B**) Distinction between $EM^{low}PM^{low}$ gliomas and control brain tissues at the whole transcriptome level identified in the REMBRANDT data set. The 65 $EM^{low}PM^{low}$ gliomas and 21 non-tumor control brain tissues in the REMBRANDT data set were compared at the whole genome level. Heat map at p = 1.0 E-7 and q = 1.08E-5 with 502 highly variably expressed probe sets is shown. Seventy-seven probe sets, corresponding to 58 genes were found over-expressed in the $EM^{low}PM^{low}$ gliomas.

(**C**) Genes over-expressed in the $EM^{low}PM^{low}$ gliomas identified in the REMBRANDT data set were up-regulated in the $EM^{low}PM^{low}$ gliomas in GSE4290 data set, but under-expressed in the non-tumor samples in both GSE4290 and GSE16011 data sets. Left panel: the 18 $EM^{low}PM^{low}$ gliomas and 23 epileptic brain (non-tumor, NT) samples with $EM^{low}PM^{low}$ phenotype in the GSE4290 data set were compared using *t*-test for the expression of genes over-expressed in the $EM^{low}PM^{low}$ gliomas in the REMBRANDT data set. Heat map at p = 0.01, q = 0.018 shows that a high fraction of these genes are up-regulated in the $EM^{low}PM^{low}$ gliomas, but under-expressed in the NT samples. Right panel: a high fraction of genes up-regulated in the $EM^{low}PM^{low}$ gliomas in the REMBRANDT data set are under-

expressed in the control NT brain samples in the GSE16011 data set. Heat map at $p = 0.01$, $q = 0.018$ is shown.

Fig. S5. Superiority of EM/PM classification scheme to morphological classification in stratifying prognosis of glioma patients in REMBRANDT, Tiantan, GSE16011 and TCGA mRNA-seq data sets.

Across all of these data sets, EM/PM classification scheme distinguished better prognostic PM or $EM^{low}PM^{low}$ gliomas from poor prognostic EM gliomas among morphological glioma subgroups which did not distinguish gliomas with distinct patients survival. The left panels show the Kaplan-Meier plots of the overall survival of the patients from morphologically diagnosed gliomas, the right panels show the Kaplan-Meier plots of the overall survival of the patients from the same gliomas as in the left panels re-classified according to EM/PM scheme.

Fig. S6. NMF clustering of glioma data sets consisting of both low-grade and high-grade gliomas with the classifiers of Verhaak *et al*. (8).

NMF clustering using TCGA classifiers was performed in the Tiantan, GSE16011 and REMBRANDT data sets. Representative heatmap of the Tiantan data set (left) and Kaplan-Meier survival curves of clusters in the Tiantan (upper right), GSE16011 (middle right) and REMBRANDT (lower right) data sets are shown. Mutually exclusive gliomas with

EGFR amplification/PTEN deletion or with co-deletion of 1p36/19q13 were found in the *NL* and *Mes* clusters.

Fig. S7. NMF clustering of glioma data sets consisting of both low-grade and high-grade gliomas with the classifiers of Phillips *et al*. (7).

NMF clustering using the classifiers of Phillips *et al*. was performed in the Tiantan (upper), GSE16011 (middle) and REMBRANDT (lower) data sets at K = 3. *dA201G10.1* was not detected in Tiantan data set, while *dA201G10.1* and *FAM20C* were not detected in GSE16011. Heatmaps of *PN*, *Prolif* and *Mes* signatures in glioma samples and Kaplan-Meier plots of the overall survival for the patients from each molecular subtype are shown. *Mes* and *Prolif* signatures overlapped with each other, or with *PN* signatures in glioma samples. Among the three data sets, the resulting glioma clusters were not associated with consistent patterns of patients´ overall survival.

Fig. S8. Comparison between EM/PM classification and the classification scheme of G-CIMP phenotype (9, 10).

Using supervised hierarchical clustering, the expression profiles of the top 100 or 200 probe sets distinguishing between G-CIMP+ and G-CIMP- gliomas were analyzed in GSE4290 and the REMBRANDT data sets in which the same probe sets could be identified. In both data sets, the probe sets enriched in G-CIMP+ gliomas were enriched in PM gliomas, whereas the probe sets enriched in G-CIMP- gliomas were enriched in EM

gliomas. Subsets of the G-CIMP+ and G-CIMP- signatures were concomitantly enriched in $EM^{low}PM^{low}$ gliomas and control samples.

Fig. S9. Subsets of the EM/PM classifiers are localized in the commonly amplified or deleted chromosomal regions in EM/PM glioma subtype-specific manner.

The normalized SNP6.0 data from the EM or PM glioma samples in the TCGA mRNA-seq data set was displayed using IGV2.0 software. The amplified or deleted chromosomal regions are color-coded in red or blue. The EM classifier PRPF31 localized at Chr 19q13.42 was amplified in 43 of 123 EM^{+++} gliomas, whereas the same region was deleted in 79 of 133 PM^{+++} gliomas. Another EM classifier NFIA localized at Chr 1p31 was deleted in 53 of 133 PM^{+++} gliomas. PM classifiers PATZ1, CSNK1E and POLR2F localized at Chr 22q was deleted in 47, 50 and 49 of 123 EM^{+++} gliomas, respectively. The red bars crossing the chromosomal scale indicate the loci of the respective classifiers. Similar findings were observed in the REMBRANDT data set (Table S8).

Fig. S10. Gene dosage-dependent expression of the EM/PM classifiers with frequent amplification or deletion.

The gene dosage effect on the expression of EM (**A**) and PM (**B**) classifiers with frequent amplification or loss was analyzed in the TCGA mRNA-seq cohort in 123 EM and 133 PM gliomas. Spearman´s rank correlation analysis on the gene expression data and SNP6.0

data-based gene dosage data was performed. The correlation coefficient (R) and p values for R are shown.

Fig. S11. Depiction of significantly amplified (red) or deleted (blue) chromosomal arm-level alterations (**A**) or chromosomal regions (**B**) of EM/PM glioma subtypes in the TCGA mRNA-seq data set along the genome.

The SNP6.0 array data from 349 gliomas in the TCGA mRNA-seq data set were analyzed using GISTIC2.0 at an amplitude threshold of ± 0.2. Loss of 1p and 19q were frequent in PM^{+++} gliomas. Gain of Chr 7, loss of Chr 10 and 22q were highly enriched in EM^{+++} gliomas. The green lines in (**B**) represent the threshold of the q value.

Fig. S12. Depiction of significantly amplified (red) or deleted (blue) chromosomal arm-level alterations (**A**) or chromosomal regions (**B**) of EM/PM glioma subtypes in the REMBRANDT data set along the genome.

The 50K *HindIII* SNP array data from 205 gliomas in the REMBRANDT data set were analyzed by GISTIC2.0 at an amplitude threshold of ± 0.2. Loss of 1p and 19q were frequent in PM^{+++} gliomas. Gain of Chr 7 and loss of Chr 10 were highly enriched in EM^{+++} gliomas, but also occurred in all other glioma subtypes at lower frequency and intensity. The green lines in (**B**) represent the threshold of the q value, known or interesting candidate genes within the regions with q value < 0.05 are indicated.

.

Fig. S13. Differential involvement of NSC compartment and glial lineage in EM/PM glioma subtypes analyzed in the Tiantan and GSE16011 data set.

In both Tiantan and GSE16011 data set, the M13C co-expression module characteristic of SVZ astrocytes was enriched in EM gliomas (**A**). Genes highly enriched in immature oligodendrocytes and immature astrocytes were enriched in PM^{+++} and EM^{+++} gliomas, respectively (**B**). Genes highly enriched in mature oligodendrocytes and mature neurons were enriched in $EM^{low}PM^{low}$ gliomas (**C**). The GSE16011 data set did not contain $EM^{low}PM^{low}$ gliomas.

Fig. S14. Construction of classifiers for developing glial cells and mature neural cell types.

(**A**) The 602 most variable genes between P1 to P8 astrocytes and OPCs and non-myelinated oligodendrocytes in GSE9566 were exported as developing astrocyte (371 genes) or developing oligodendrocyte (231 genes) classifiers. Two-group comparison at a FDR of 3.2×10^{-4} was used. Gene lists will be provided upon request.

(**B**) The transcriptome data of the mature astrocytes, oligodendrocytes and neurons in GSE9566 were analyzed with unsupervised hierarchical clustering after variance filtering, excluding genes with standard deviations below 48% of the maximal standard deviation. Highly variably expressed genes for each cell type were exported as classifiers. Gene lists will be provided upon request.

(**C**) Dendrogram and sample clustering of purified mouse CNS cell types (Cahoy J.D. *et al*. J. Neurosci. 2008. 28:264 –278) used to generate gene lists in (**A**) and (**B**).

List of Supplementary Tables

Table S1. Glioma data sets analyzed in this study.

Table S2. Members of EM and PM classifiers.

Table S3. Clinical and genomic characteristics of EM/PM glioma subtypes.

Table S4. Genes highly enriched in EM/PM glioma subtypes.

Table S5. Hazard ratios for association between EM/PM scores and the overall mortality using multivariate Cox regression analysis

Table S6. Cross-labeling of glioma samples in NMF clustering by TCGA or EM/PM classifiers.

Table S7. Subset of EM/PM classifiers are concomitantly gained or lost in glioma genomes in EM/PM glioma subtype-specific manner. Data are derived from the TCGA mRNA-seq data set.

Table S8. Subset of EM/PM classifiers are concomitantly gained or lost in glioma genomes in EM/PM subtype-specific manner. Data are derived from the REMBRANDT data set.

Table S9. Co-occurrence of SCNAs of EM/PM classifiers in EM and PM gliomas of the TCGA mRNA-seq data set.

Table S10. Contribution of the gene dosage-dependent expression of individual EM/PM classifiers to the EM/PM signature in glioma samples.

Table S11. Regional chromosomal alterations in PM^{+++} and EM^{+++} gliomas from the TCGA mRNA-seq data set.

Table S12. Regional chromosomal alterations in EM/PM gliomas from the REMBRANDT data set.

Table S13. GSEA analysis for differential enrichment of developing astrocyte or oligodendrocyte signatures in EM/PM gliomas.

Supplementary References

1. Sun L, *et al*. (2006) Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer cell* 9(4):287-300.
2. Gravendeel LA, *et al*. (2009) Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer research* 69(23):9065-9072.
3. Madhavan S, *et al*. (2009) Rembrandt: helping personalized medicine become a reality through integrative translational research. *Mol Cancer Res* 7(2):157-167.
4. Cahoy JD, *et al*. (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* 28(1):264-278.
5. Louis DN, *et al*. (2007) The 2007 WHO classification of tumours of the central nervous system. *Acta neuropathologica* 114(2):97-109.
6. Brunet JP, Tamayo P, Golub TR, & Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci USA* 101(12):4164-4169.
7. Phillips HS, *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(3):157-173.

8. Verhaak RG, *et al*. (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell* 17(1):98-110.

9. Baysan M, *et al*. (2012) G-cimp status prediction of glioblastoma samples using mRNA expression data. *PLoS ONE* 7(11):e47839.

10. Noushmehr H, *et al*. (2010) Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer cell* 17(5):510-522.

11. Oldham MC, *et al*. (2008) Functional organization of the transcriptome in human brain. *Nat Neurosci* 11(11):1271-1282.

12. Subramanian A, *et al*. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102(43):15545-15550.

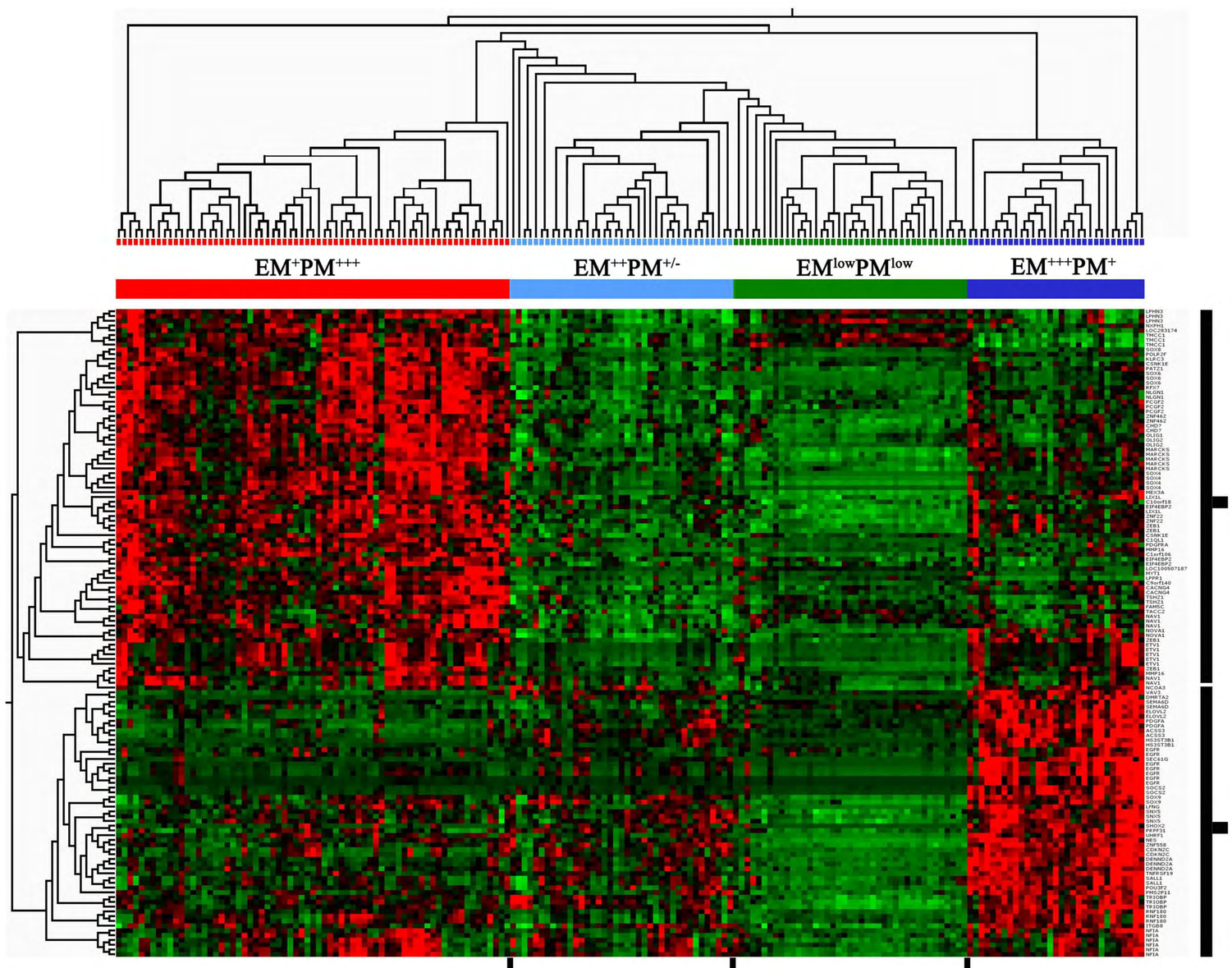
13. Lin M, *et al*. (2004) dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* 20(8):1233-1240.

14. Hupe P, Stransky N, Thiery JP, Radvanyi F, & Barillot E (2004) Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics* 20(18):3413-3422.

15. Mermel CH, *et al*. (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology* 12(4):R41.

Fig. S1

GSE4290 (N = 180)

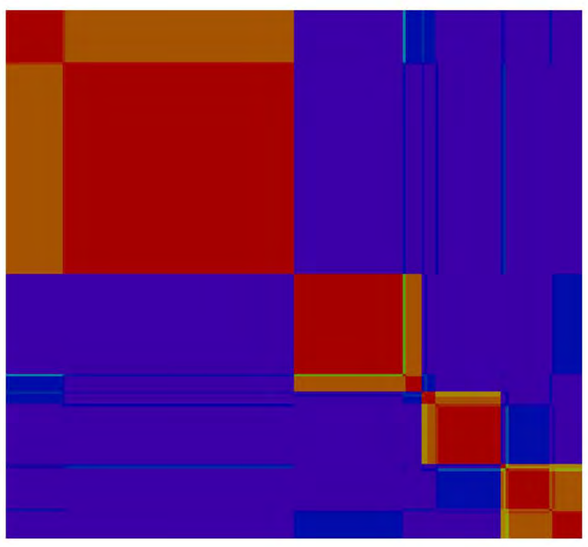


	EM+PM+++	EM++PM+/-	EMlowPMlow	EM+++PM+
N	69	39	41	31
Astrocytoma II	3	3	1	0
Astrocytoma III	10	6	1	2
GBM	18	26	7	26
Oligodendroglioma II	30	1	7	0
Oligodendroglioma III	7	2	1	2
Unknown	1	1	1	1
Non-tumor	0	0	23	0

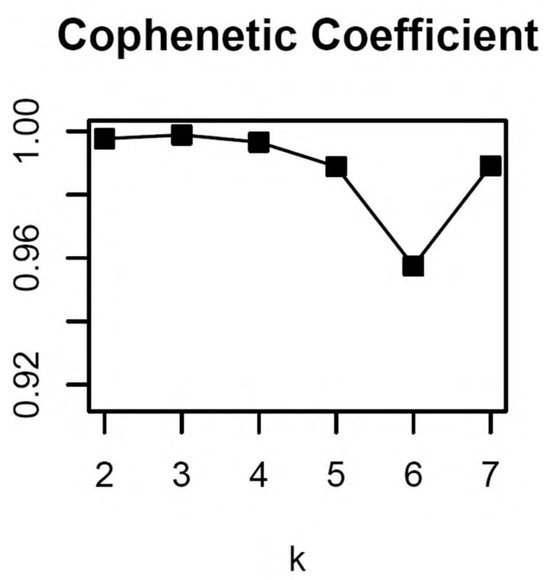
PDGFRA	*OLIG1*	*OLIG2*	*SOX8*
SOX6	*ZEB1*	*SOX4*	*MYT1*
PCGF2	*CHD7*	*POLR2F*	*TSHZ1*
ETV1	*CSNK1E*	*MARCKS*	*TACC2*
PATZ1	*LPHN3*	*KLRC3*	*NLGN1*
LPPR1	*NAV1*	*NXPH1*	*CACNG4*
EIF4EBP2	*NOVA1*	*MMP16*	*C9orf140*
LIX1L	*MEX3A*	*FAM5C*	*C1orf106*
ZNF22	*TMCC1*	*C10orf18*	*LOC283174*
C1QL1	*LOC100507187*	*RFX7*	*ZNF462*

EGFR	*SEC61G*	*VAV3*	*CDKN2C*
TNFRSF19	*NES*	*PDGFA*	*SOX9*
POU3F2	*UHRF1*	*NFIA*	*LFNG*
TRIOBP	*ACSS3*	*ELOVL2*	*HS3ST3B1*
SALL1	*PRPF31*	*ITGB8*	*SEMA6D*
NCOA3	*SNX5*	*RNF180*	*DMRTA2*
SOCS2	*SHOX2*	*ZNF558*	*PMS2P11*
DENND2A			

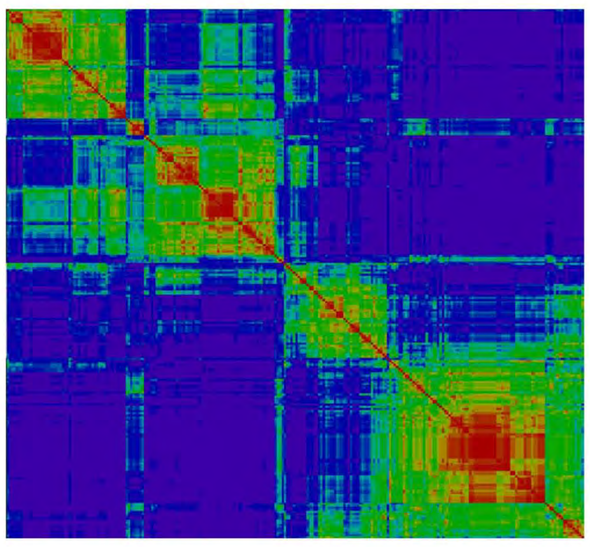
Tiantan (k = 4)



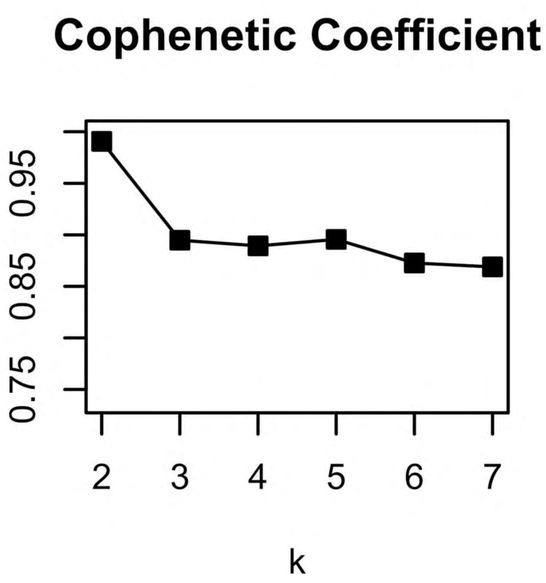
Cophenetic coef.=0.9966



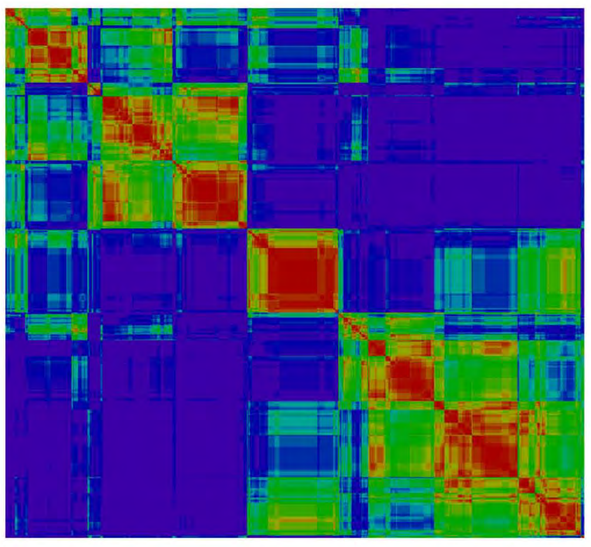
GSE16011 (k = 5)



Cophenetic coef.=0.8956



REMBRANDT (k = 4)



Cophenetic coef.=0.877

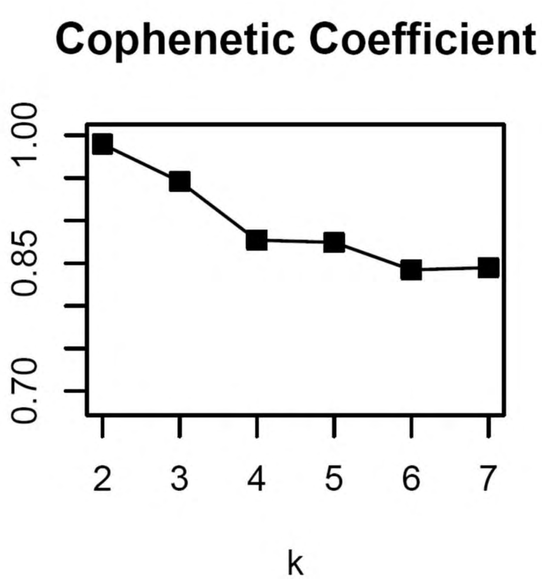
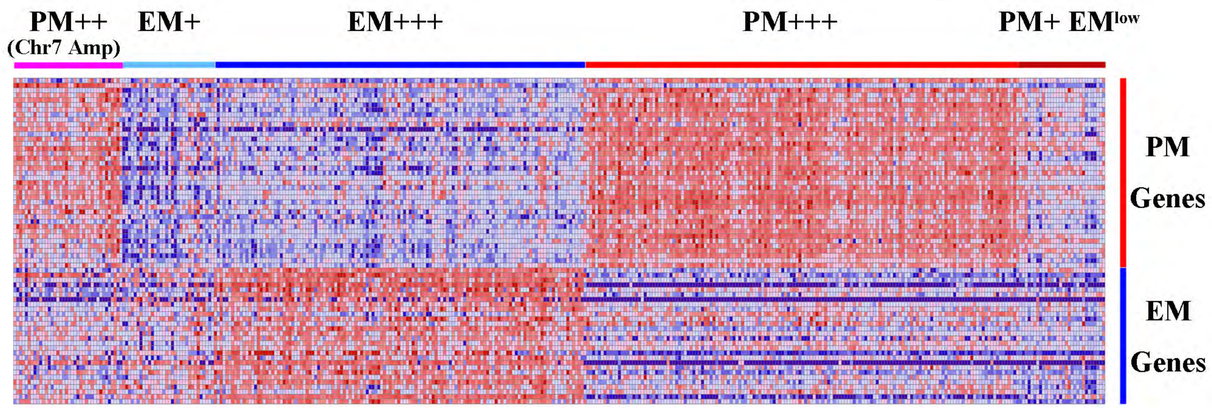
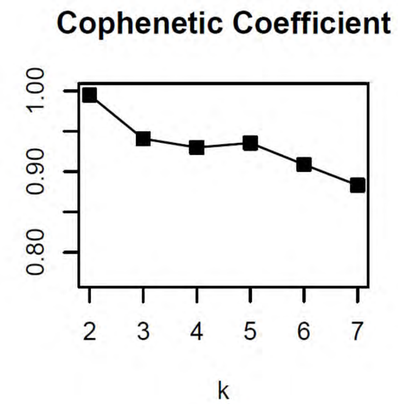


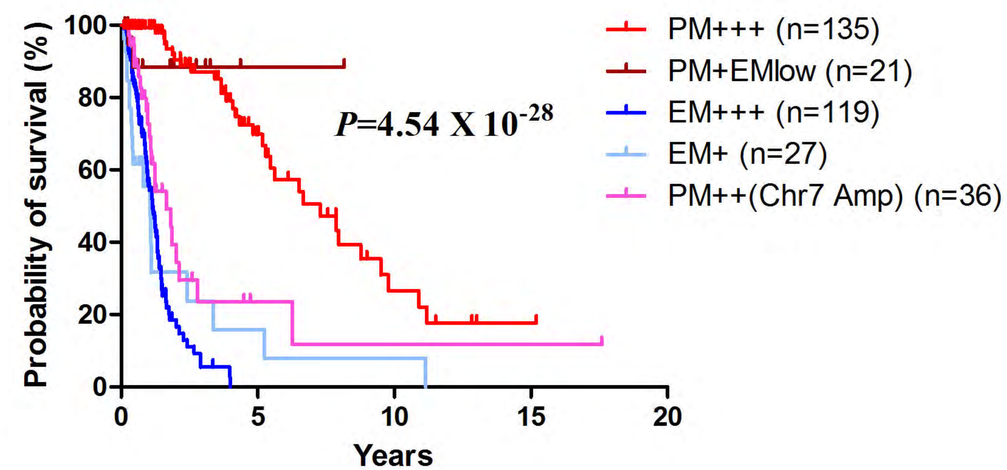
Fig. S2

TCGA mRNA-seq cohort (N=381, k=5)





	PM++ (Chr7 Amp)	EM+	EM+++	PM+++	PM+ EMlow	
Number of samples	38	32	129	155	27	
A II	0	2	0	12	3	
A III	2	1	16	24	2	
O II	0	0	1	42	9	
O III	1	4	1	27	4	
OA II	0	0	1	20	4	
OA III	3	0	4	19	3	
GBM IV	24	19	78	0	0	
Unknown	8	6	28	11	2	
Age at diagnosis (Median ± SD)	50 ± 16.5	55 ± 15.7	59 ± 10.8	39 ± 12.8	38 ± 13.7	$p<0.0001$
Survival Years (Median ± SD)	1.1 ± 3.3	0.5 ± 0.7	0.9 ± 0.7	1.4 ± 3.2	1.8 ± 2.0	$p<0.0001$
IDH1 mutation	11/35	2/29	0	139/147	16/25	
1p36/19q13 co-deletion	5/36	0	6/123	53/133	10/26	
EGFR Amp/Mut PTEN Del/Mut	28/36	23/31	126/126	24/133	4/26	
Focal Amp EGFR	6/36	2/31	83/123	3/133	0	
Focal Amp PDGFRA	10/36	4/31	13/123	5/133	1	
Co-Amp EGFR/PDGFRA	2/36	0	10/123	0	0	



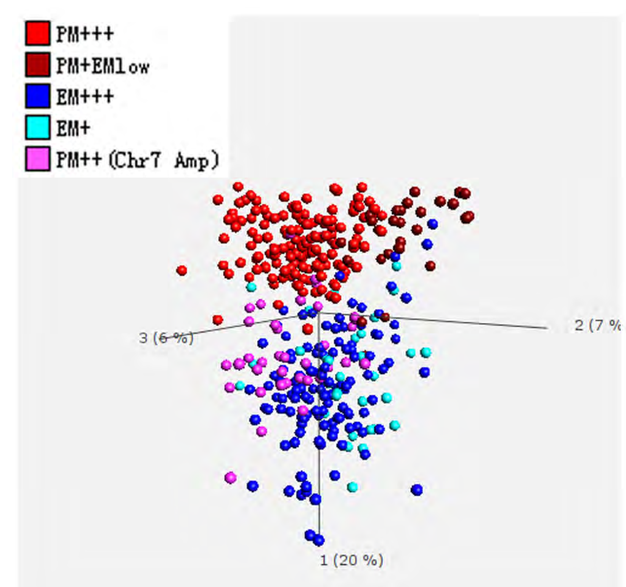


Fig. S3

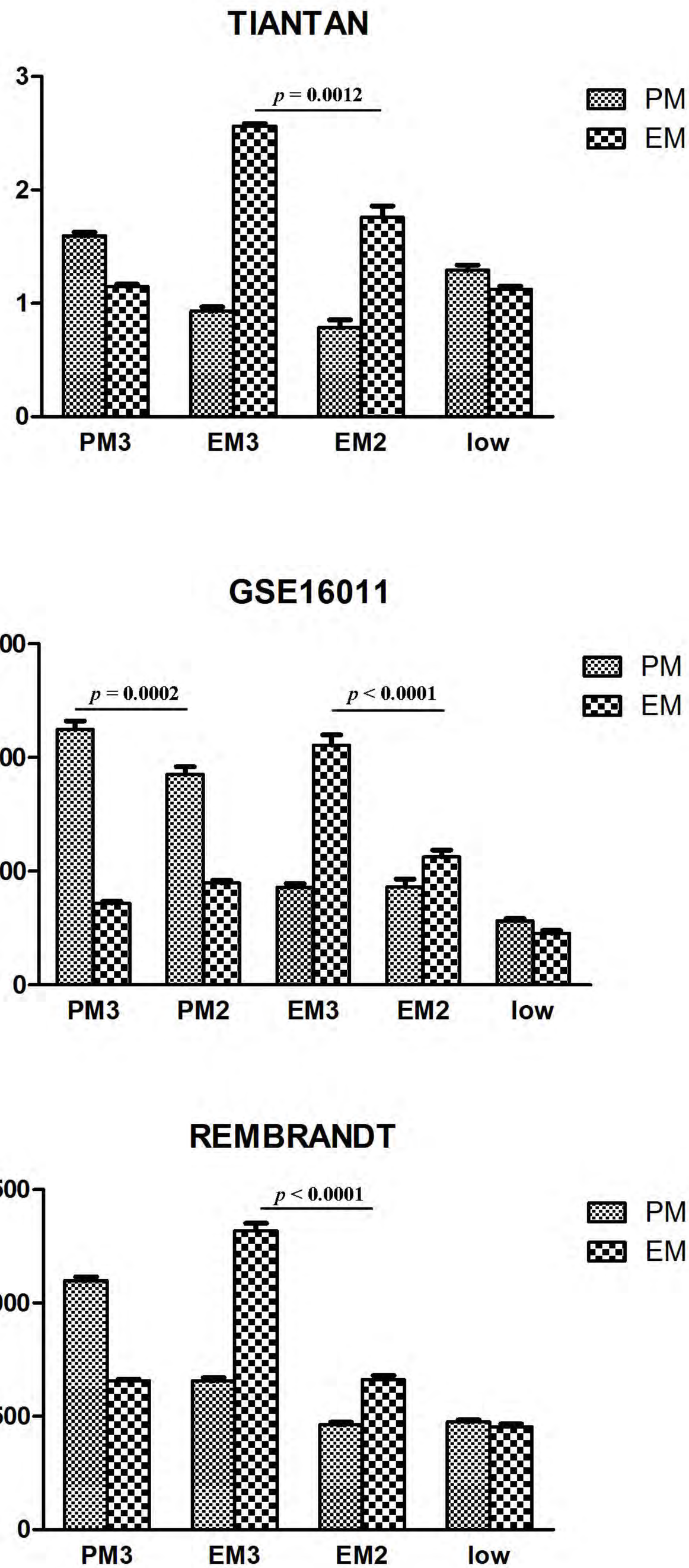

TIANTAN
p = 0.0012
PM
EM
3
2
1
0
PM3
EM3
EM2
low
GSE16011
p = 0.0002
p < 0.0001
PM
EM
1500
1000
500
0
PM3
PM2
EM3
EM2
low
REMBRANDT
p < 0.0001
PM
EM
1500
1000
500
0
PM3
EM3
EM2
low

Fig. S4

A

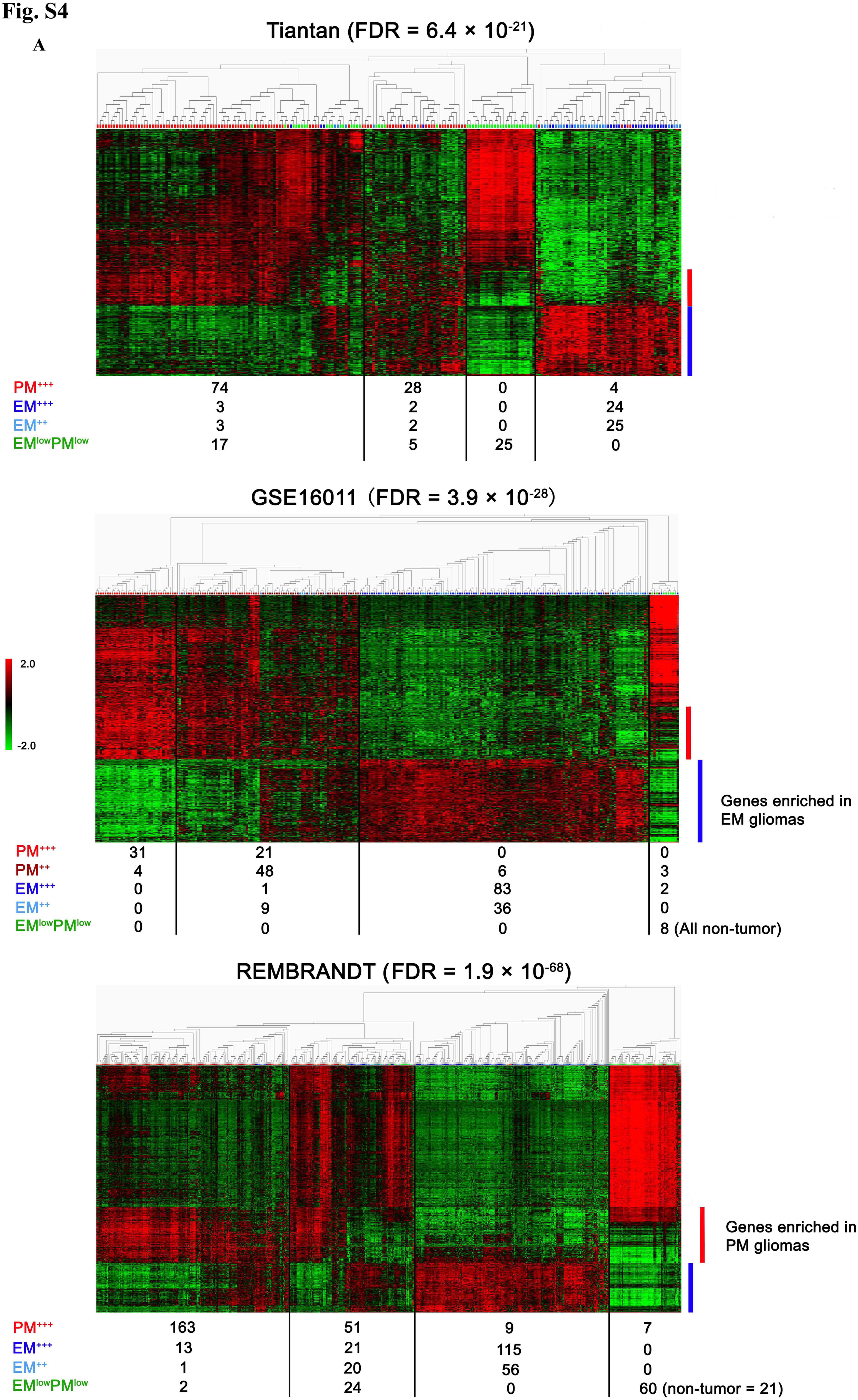
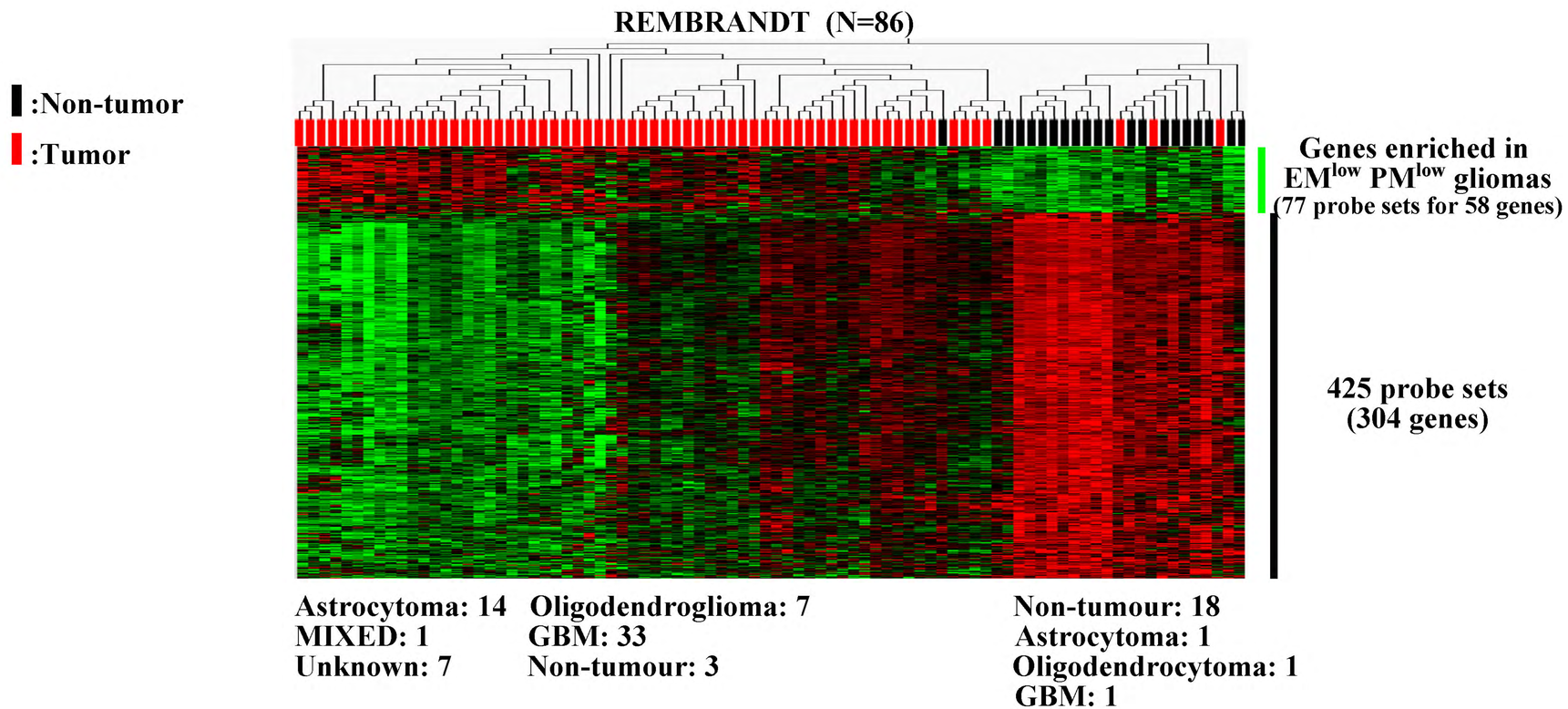

Tiantan (FDR = 6.4 × 10-21)
PM+++ 74 28 0 4
EM+++ 3 2 0 24
EM++ 3 2 0 25
EMlowPMlow 17 5 25 0
GSE16011 (FDR = 3.9 × 10-28)
2.0
-2.0
PM+++ 31 21 0 0
PM++ 4 48 6 3
EM+++ 0 1 83 2
EM++ 0 9 36 0
EMlowPMlow 0 0 0 8 (All non-tumor)
Genes enriched in EM gliomas
REMBRANDT (FDR = 1.9 × 10-68)
PM+++ 163 51 9 7
EM+++ 13 21 115 0
EM++ 1 20 56 0
EMlowPMlow 2 24 0 60 (non-tumor = 21)
Genes enriched in PM gliomas

Fig. S4

B


REMBRANDT (N=86)
:Non-tumor
:Tumor
Genes enriched in
EMlow PMlow gliomas
(77 probe sets for 58 genes)
425 probe sets
(304 genes)
Astrocytoma: 14
MIXED: 1
Unknown: 7
Oligodendroglioma: 7
GBM: 33
Non-tumour: 3
Non-tumour: 18
Astrocytoma: 1
Oligodendrocytoma: 1
GBM: 1

C

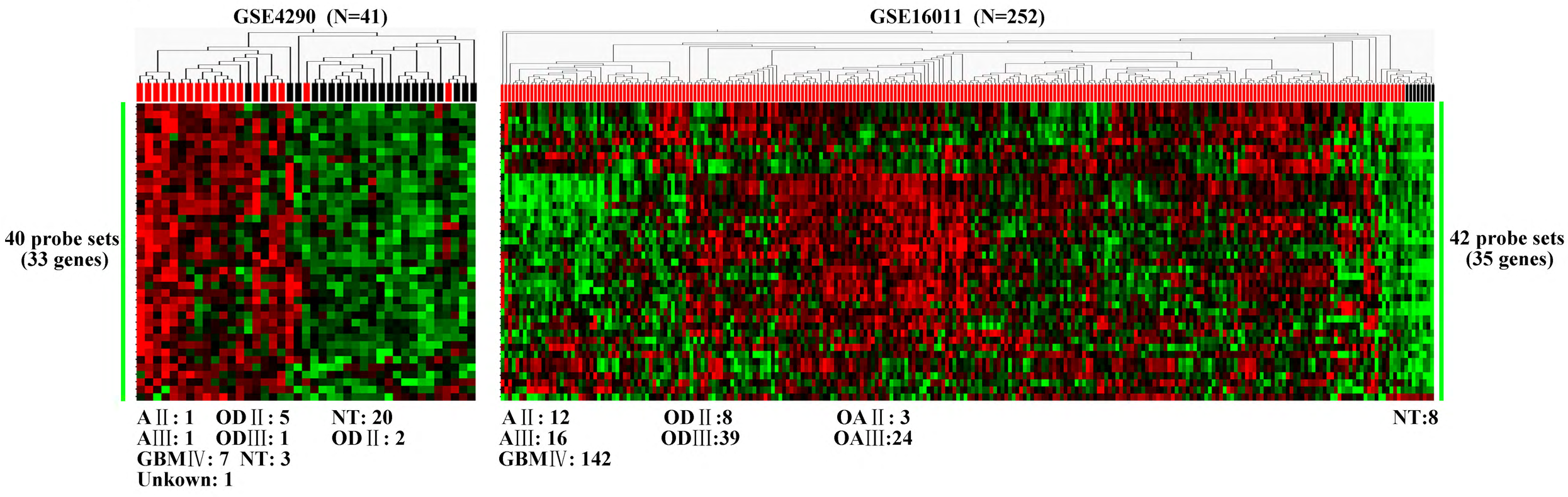

GSE4290 (N=41)
GSE16011 (N=252)
40 probe sets
(33 genes)
42 probe sets
(35 genes)
AⅡ: 1
AⅢ: 1
GBMⅣ: 7
Unkown: 1
ODⅡ: 5
ODⅢ: 1
NT: 3
NT: 20
ODⅡ: 2
AⅡ: 12
AⅢ: 16
GBMⅣ: 142
ODⅡ:8
ODⅢ:39
OAⅡ: 3
OAⅢ:24
NT:8

Fig. S5

REMBRANDT

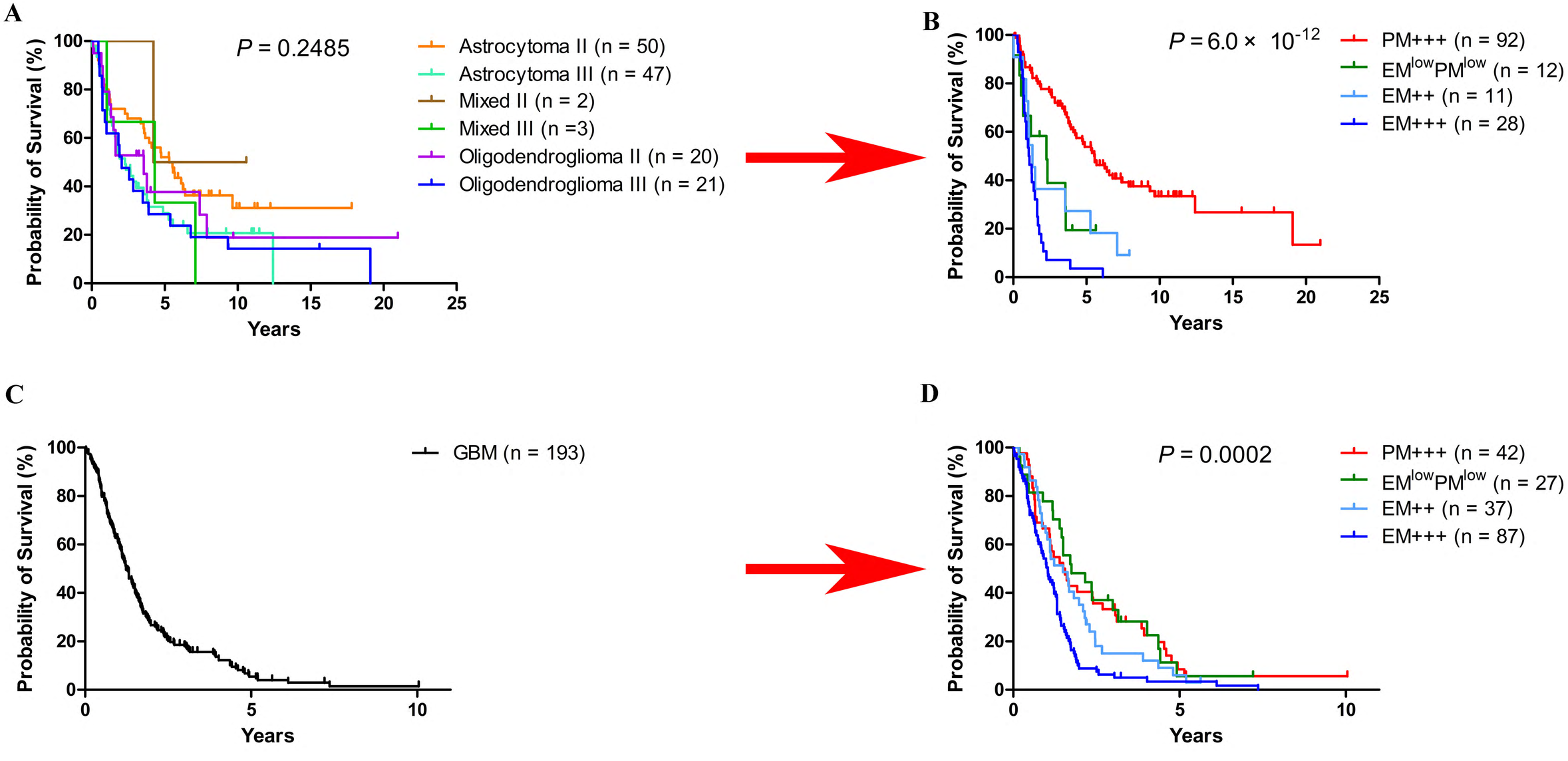


Fig. S5

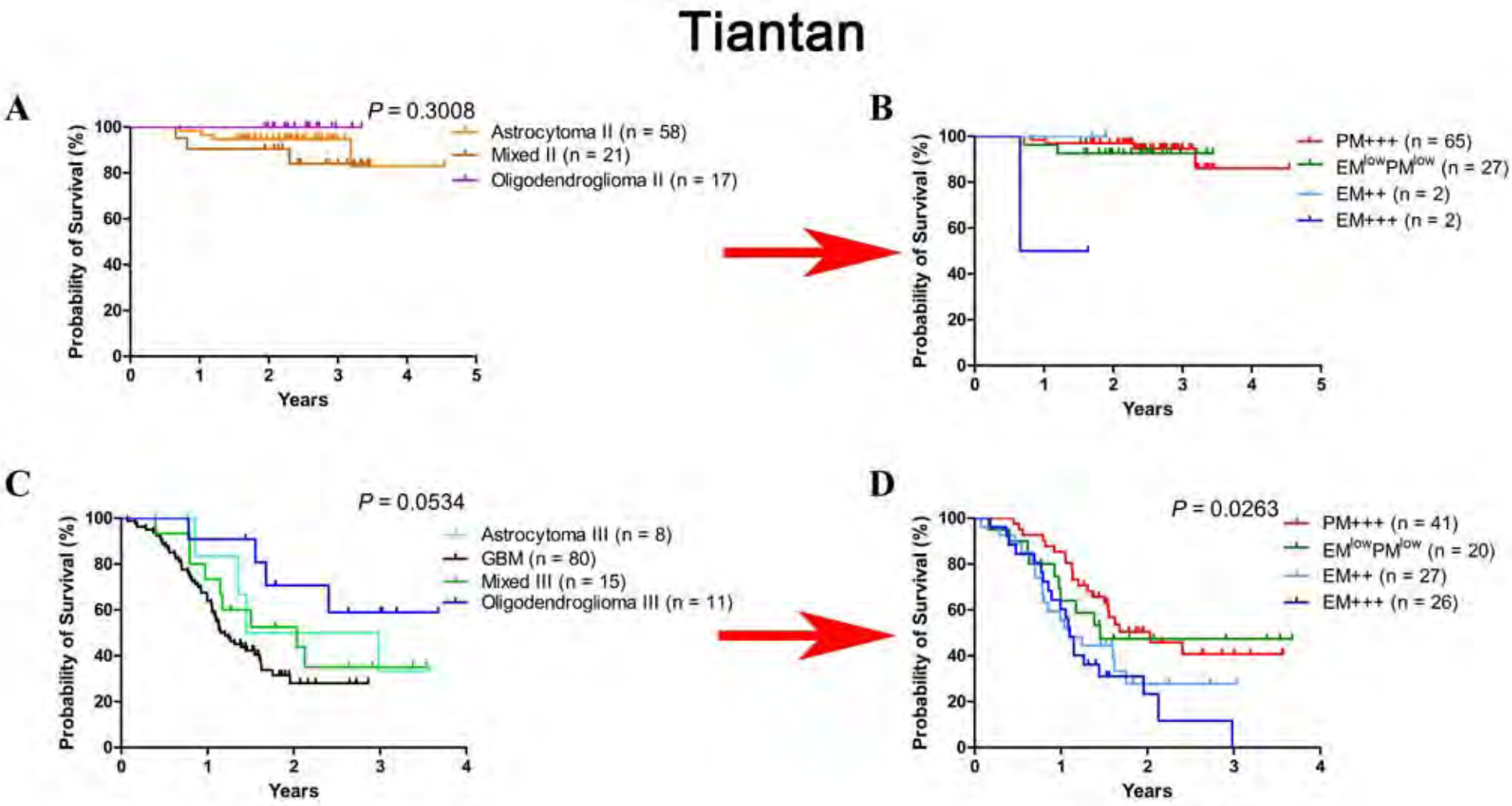

Tiantan
A
P = 0.3008
Astrocytoma II (n = 58)
Mixed II (n = 21)
Oligodendroglioma II (n = 17)
Probability of Survival (%)
Years
B
PM+++ (n = 65)
EMlowPMlow (n = 27)
EM++ (n = 2)
EM+++ (n = 2)
Probability of Survival (%)
Years
C
P = 0.0534
Astrocytoma III (n = 8)
GBM (n = 80)
Mixed III (n = 15)
Oligodendroglioma III (n = 11)
Probability of Survival (%)
Years
D
P = 0.0263
PM+++ (n = 41)
EMlowPMlow (n = 20)
EM++ (n = 27)
EM+++ (n = 26)
Probability of Survival (%)
Years

Fig. S5

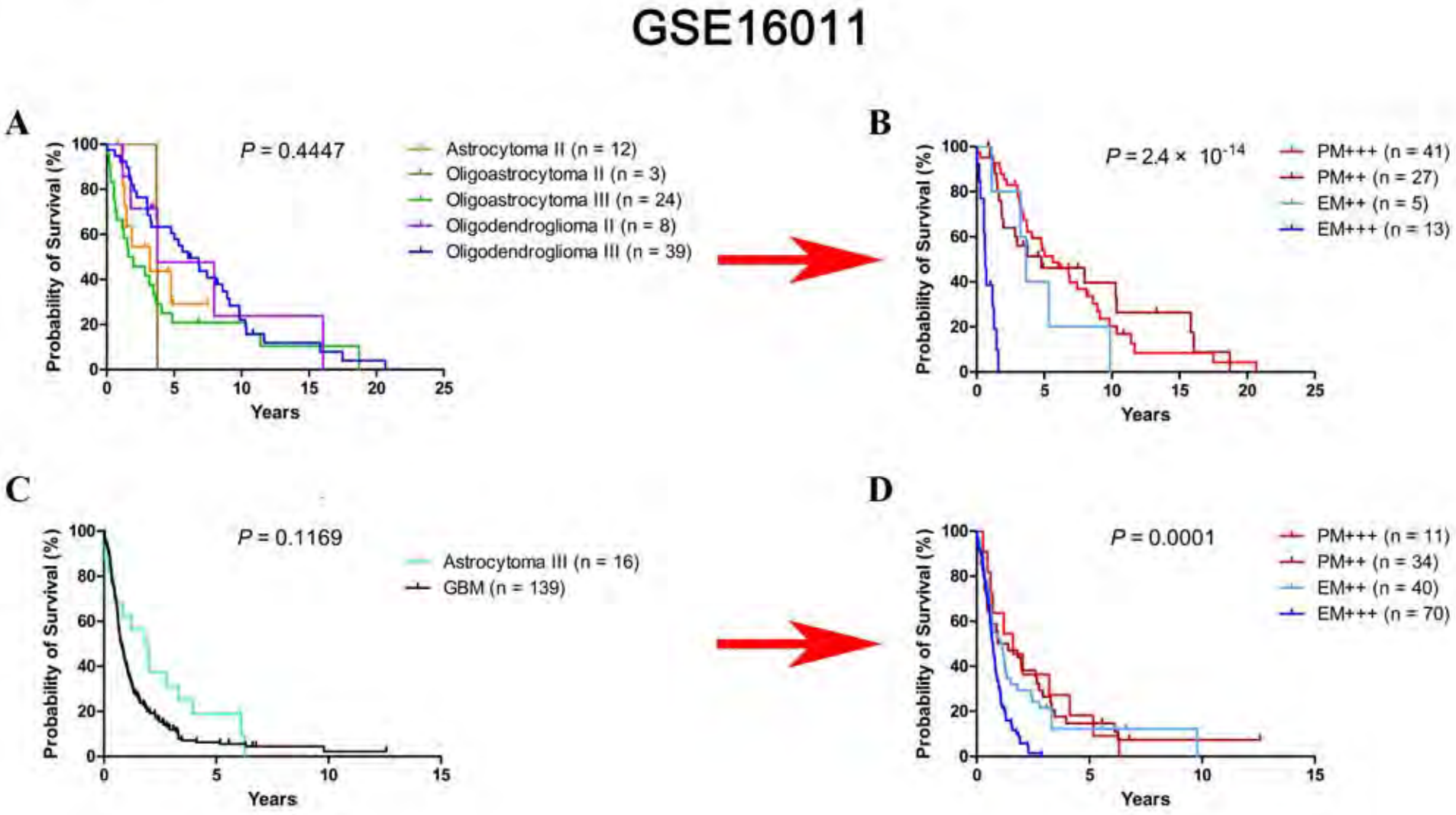

GSE16011
A
P = 0.4447
Astrocytoma II (n = 12)
Oligoastrocytoma II (n = 3)
Oligoastrocytoma III (n = 24)
Oligodendroglioma II (n = 8)
Oligodendroglioma III (n = 39)
Probability of Survival (%)
Years
B
P = 2.4 × 10-14
PM+++ (n = 41)
PM++ (n = 27)
EM++ (n = 5)
EM+++ (n = 13)
Probability of Survival (%)
Years
C
P = 0.1169
Astrocytoma III (n = 16)
GBM (n = 139)
Probability of Survival (%)
Years
D
P = 0.0001
PM+++ (n = 11)
PM++ (n = 34)
EM++ (n = 40)
EM+++ (n = 70)
Probability of Survival (%)
Years

Fig. S5

TCGA mRNA-seq

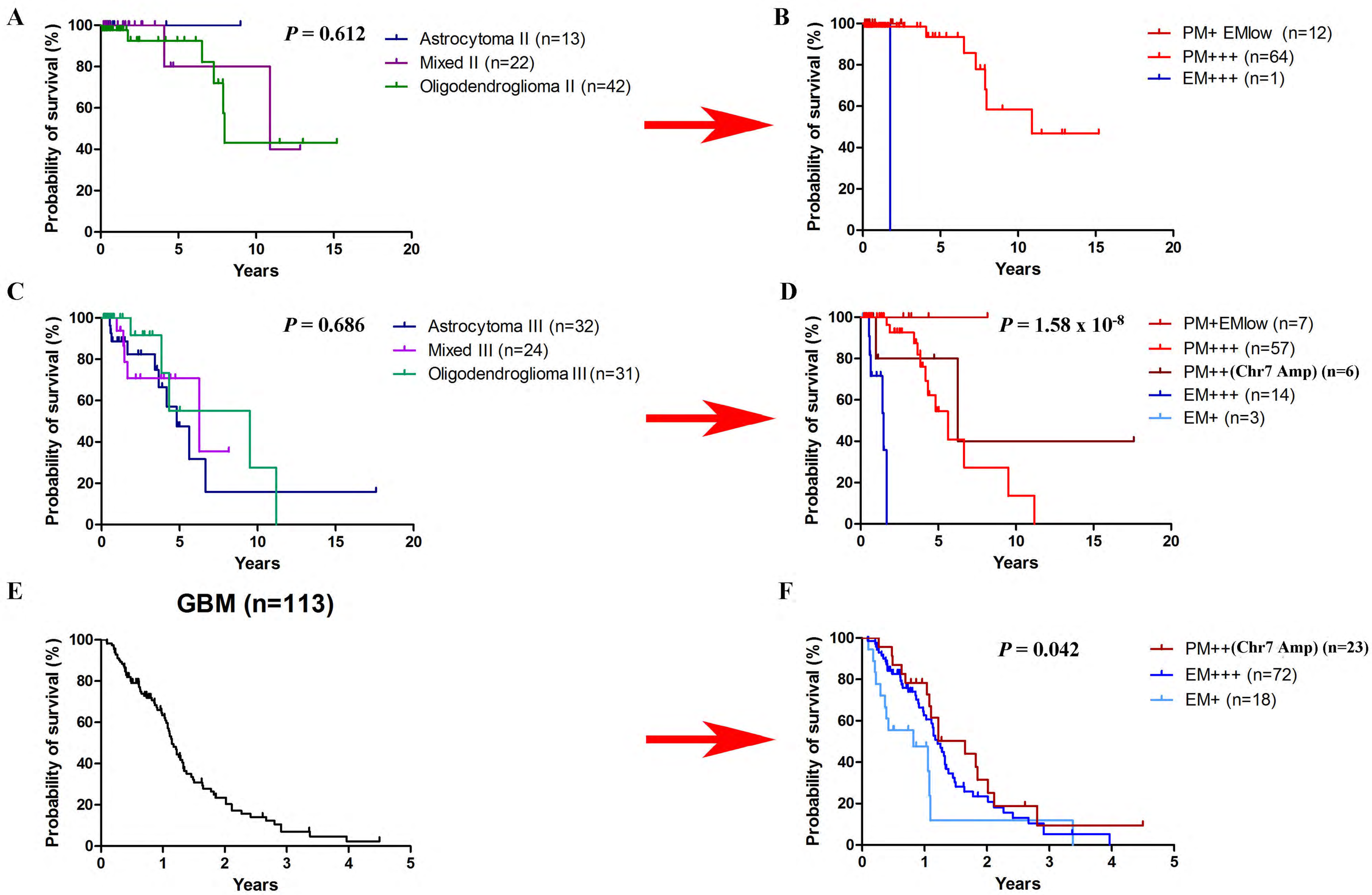


Fig. S6

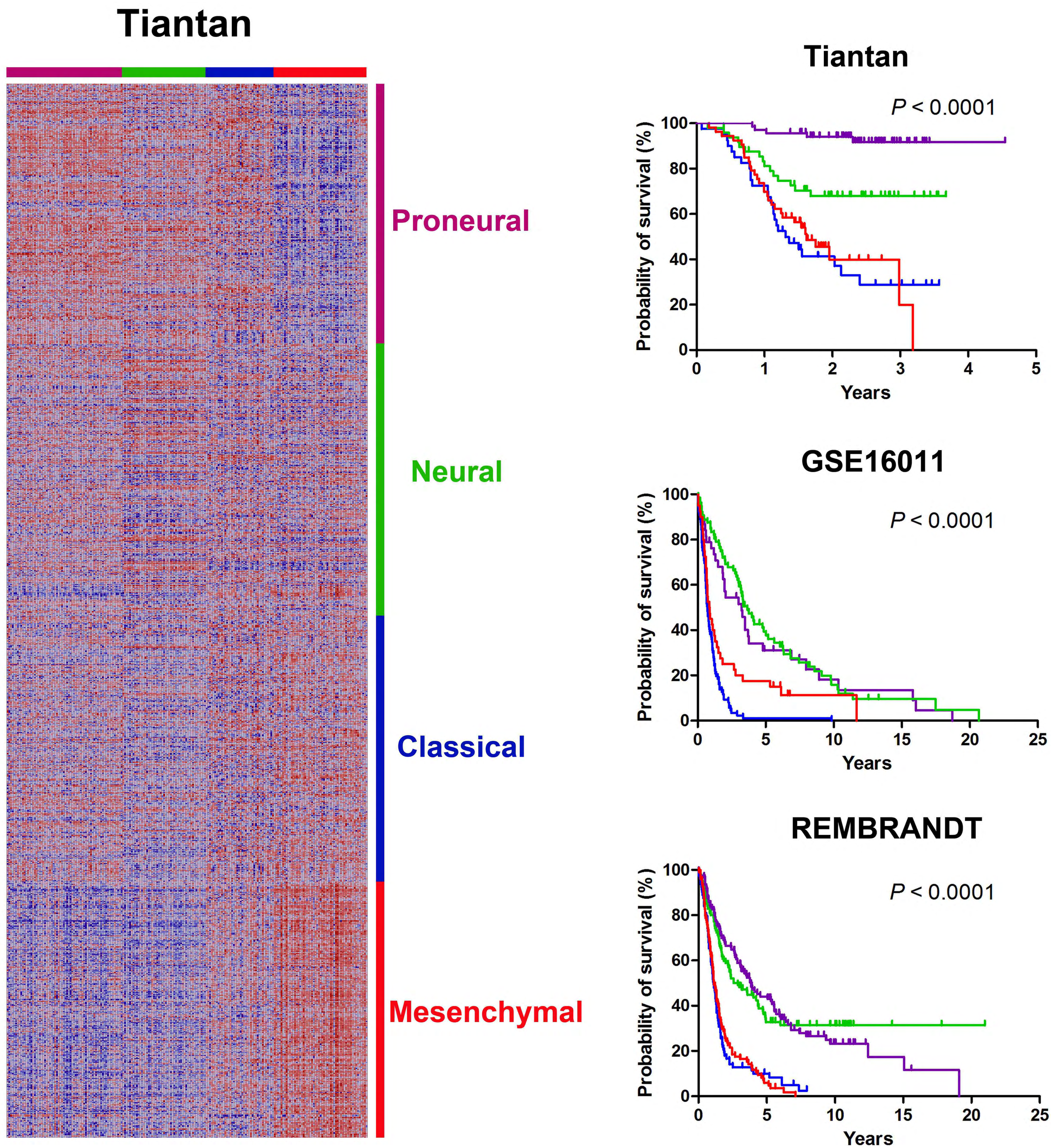

Tiantan
Proneural
Neural
Classical
Mesenchymal
Tiantan
P < 0.0001
Probability of survival (%)
Years
GSE16011
P < 0.0001
Probability of survival (%)
Years
REMBRANDT
P < 0.0001
Probability of survival (%)
Years

Fig. S7

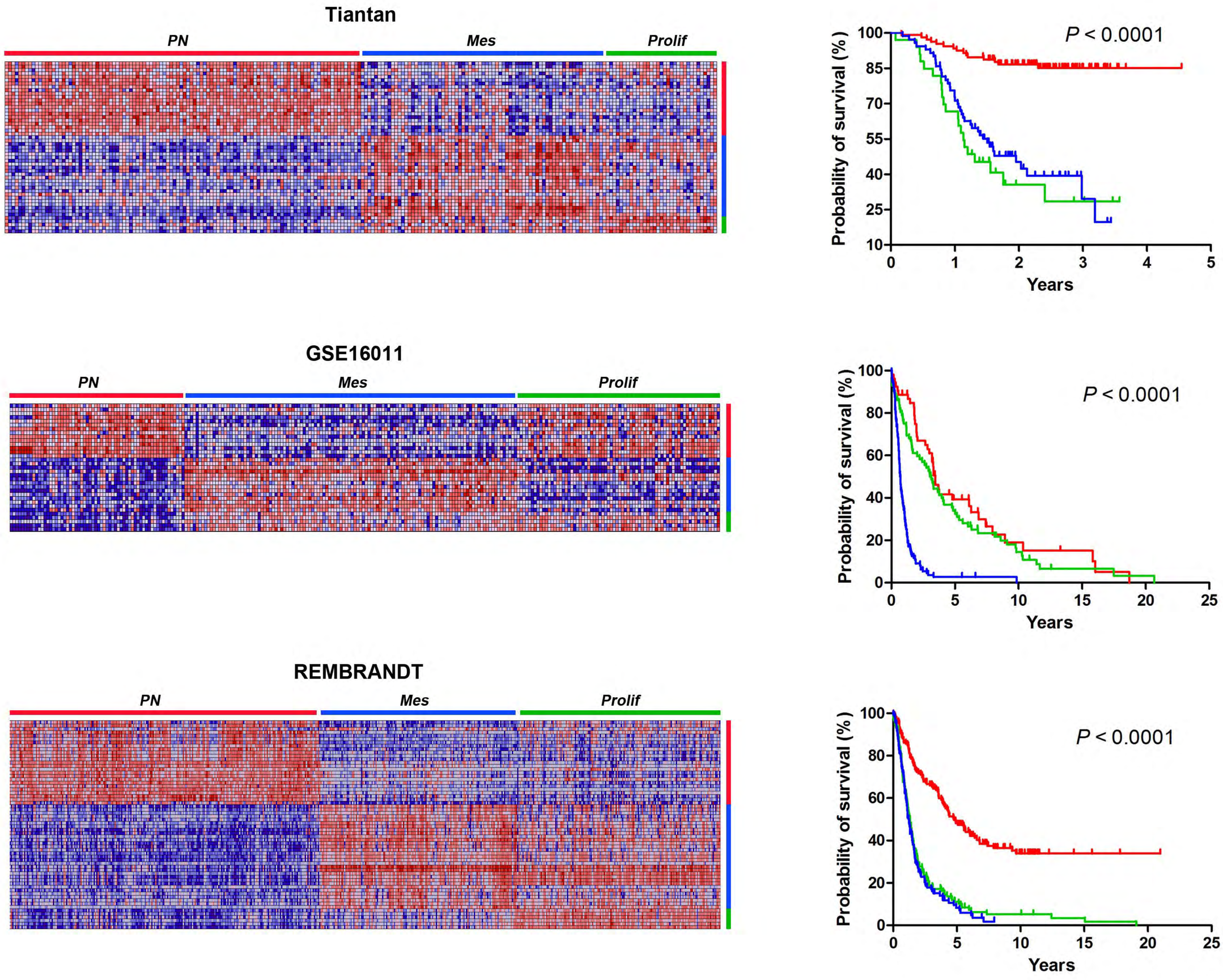

Tiantan
PN
Mes
Prolif
P < 0.0001
Probability of survival (%)
100
85
70
55
40
25
10
0
1
2
3
4
5
Years
GSE16011
PN
Mes
Prolif
P < 0.0001
Probability of survival (%)
100
80
60
40
20
0
5
10
15
20
25
Years
REMBRANDT
PN
Mes
Prolif
P < 0.0001
Probability of survival (%)
100
80
60
40
20
0
5
10
15
20
25
Years

Fig. S8

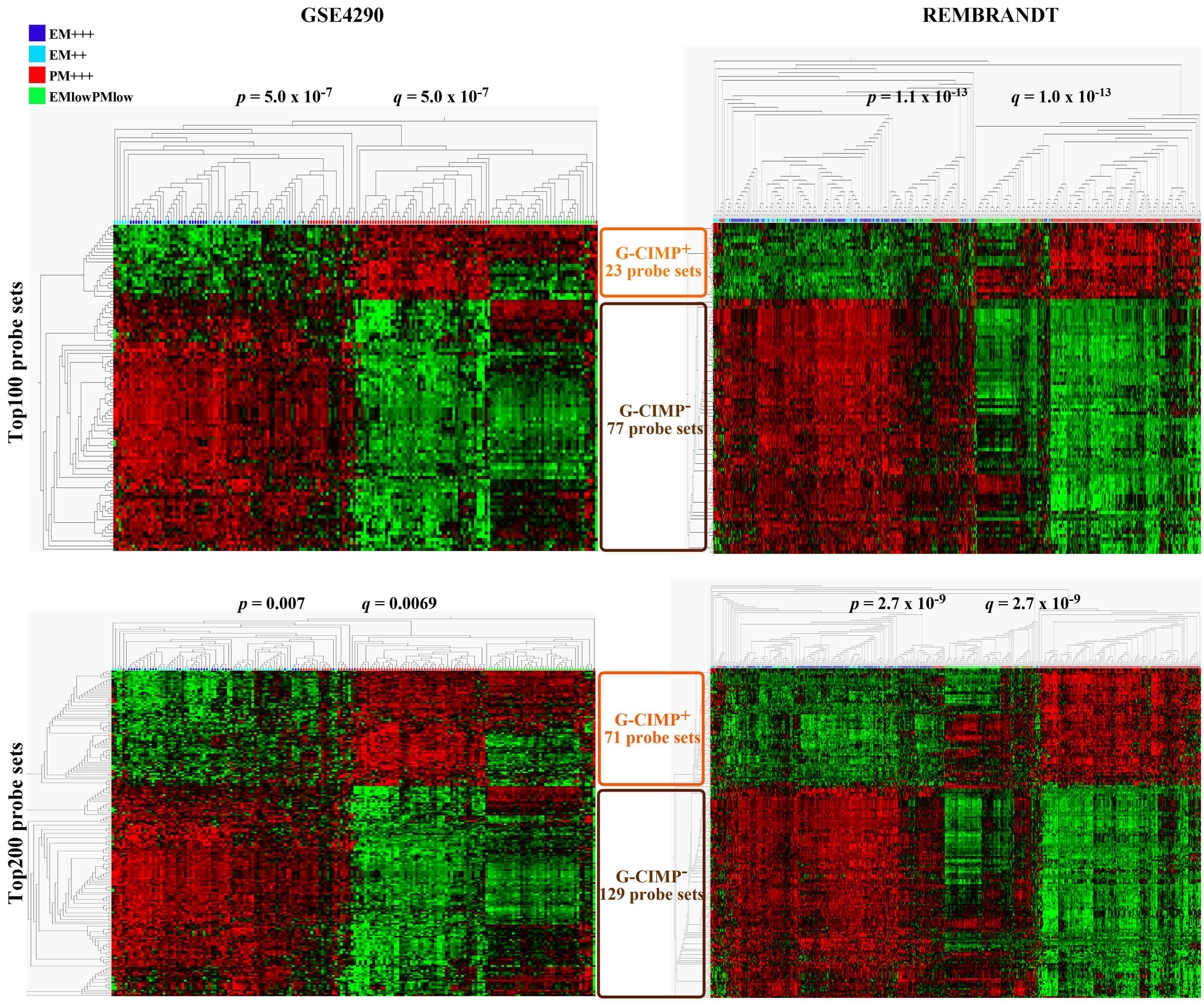

GSE4290
REMBRANDT
EM+++
EM++
PM+++
EMlowPMlow
p = 5.0 x 10-7
q = 5.0 x 10-7
p = 1.1 x 10-13
q = 1.0 x 10-13
Top100 probe sets
G-CIMP+
23 probe sets
G-CIMP-
77 probe sets
p = 0.007
q = 0.0069
p = 2.7 x 10-9
q = 2.7 x 10-9
Top200 probe sets
G-CIMP+
71 probe sets
G-CIMP-
129 probe sets

Fig. S9

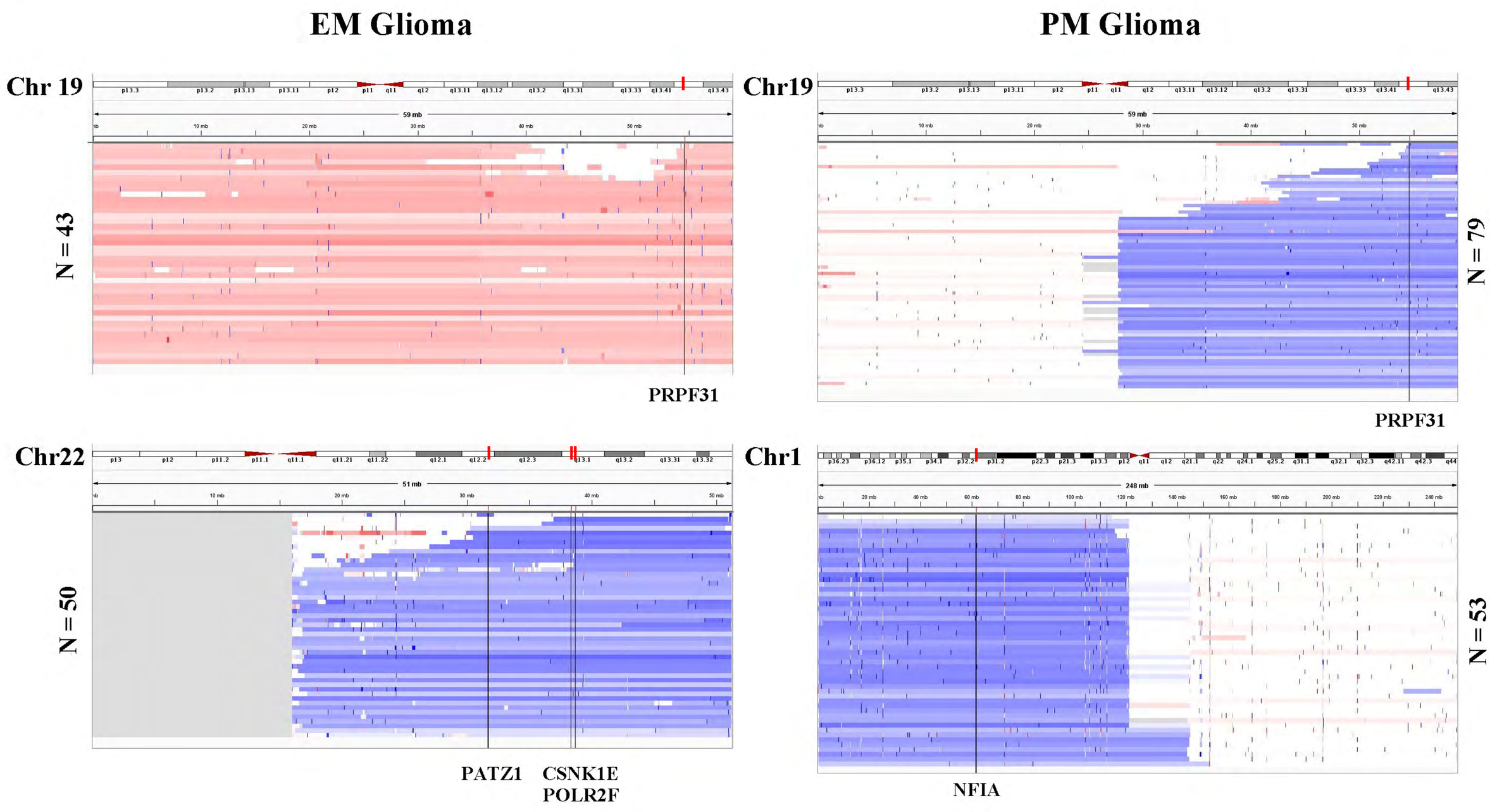

EM Glioma
PM Glioma
Chr 19
Chr19
N = 43
N = 79
PRPF31
Chr22
Chr1
N = 50
N = 53
PATZ1
CSNK1E
POLR2F
NFIA

Fig. S10

A

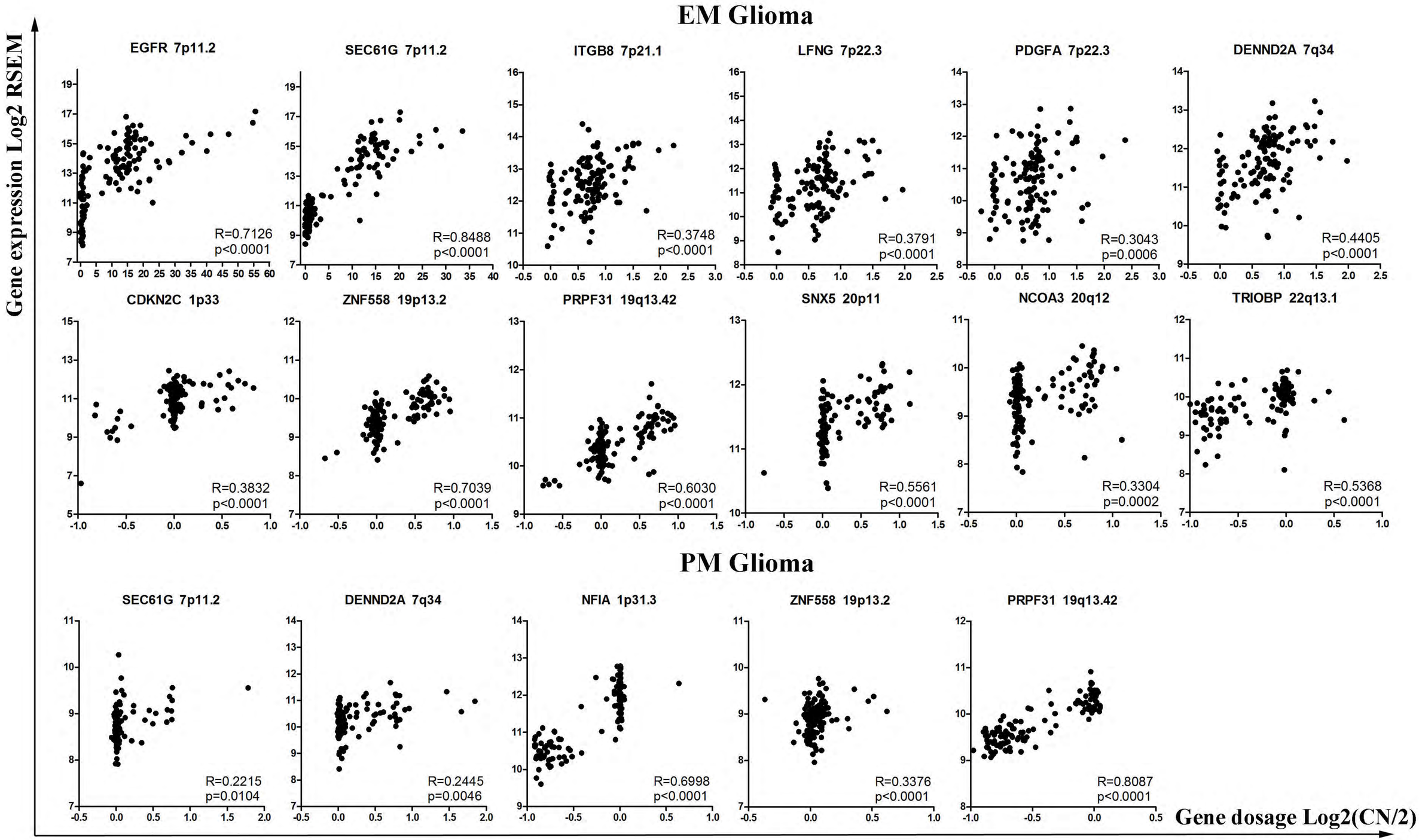

EM Glioma
Gene expression Log2 RSEM
EGFR 7p11.2
R=0.7126
p<0.0001
SEC61G 7p11.2
R=0.8488
p<0.0001
ITGB8 7p21.1
R=0.3748
p<0.0001
LFNG 7p22.3
R=0.3791
p<0.0001
PDGFA 7p22.3
R=0.3043
p=0.0006
DENND2A 7q34
R=0.4405
p<0.0001
CDKN2C 1p33
R=0.3832
p<0.0001
ZNF558 19p13.2
R=0.7039
p<0.0001
PRPF31 19q13.42
R=0.6030
p<0.0001
SNX5 20p11
R=0.5561
p<0.0001
NCOA3 20q12
R=0.3304
p=0.0002
TRIOBP 22q13.1
R=0.5368
p<0.0001
PM Glioma
SEC61G 7p11.2
R=0.2215
p=0.0104
DENND2A 7q34
R=0.2445
p=0.0046
NFIA 1p31.3
R=0.6998
p<0.0001
ZNF558 19p13.2
R=0.3376
p<0.0001
PRPF31 19q13.42
R=0.8087
p<0.0001
Gene dosage Log2(CN/2)

Fig. S10

B

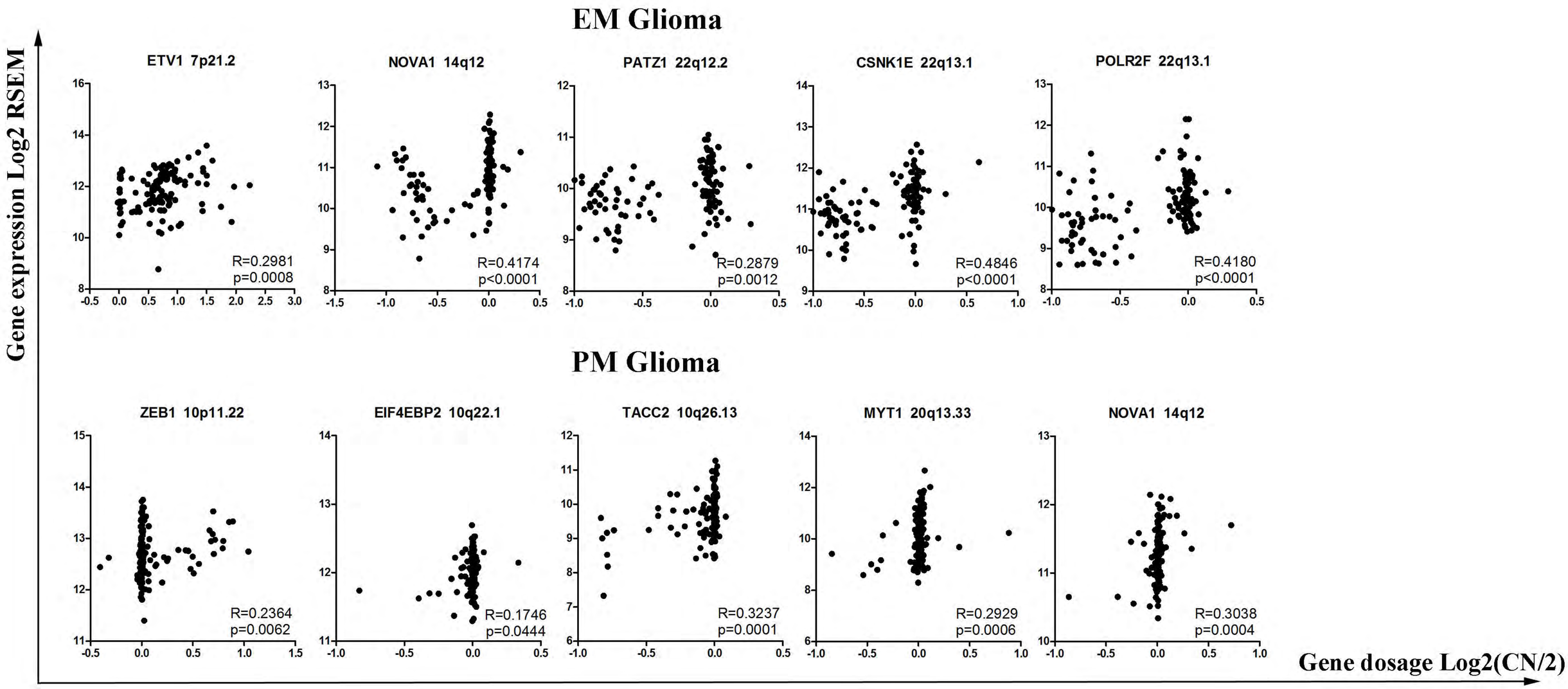

EM Glioma
ETV1 7p21.2
R=0.2981
p=0.0008
NOVA1 14q12
R=0.4174
p<0.0001
PATZ1 22q12.2
R=0.2879
p=0.0012
CSNK1E 22q13.1
R=0.4846
p<0.0001
POLR2F 22q13.1
R=0.4180
p<0.0001
PM Glioma
ZEB1 10p11.22
R=0.2364
p=0.0062
EIF4EBP2 10q22.1
R=0.1746
p=0.0444
TACC2 10q26.13
R=0.3237
p=0.0001
MYT1 20q13.33
R=0.2929
p=0.0006
NOVA1 14q12
R=0.3038
p=0.0004
Gene expression Log2 RSEM
Gene dosage Log2(CN/2)

Fig. S11

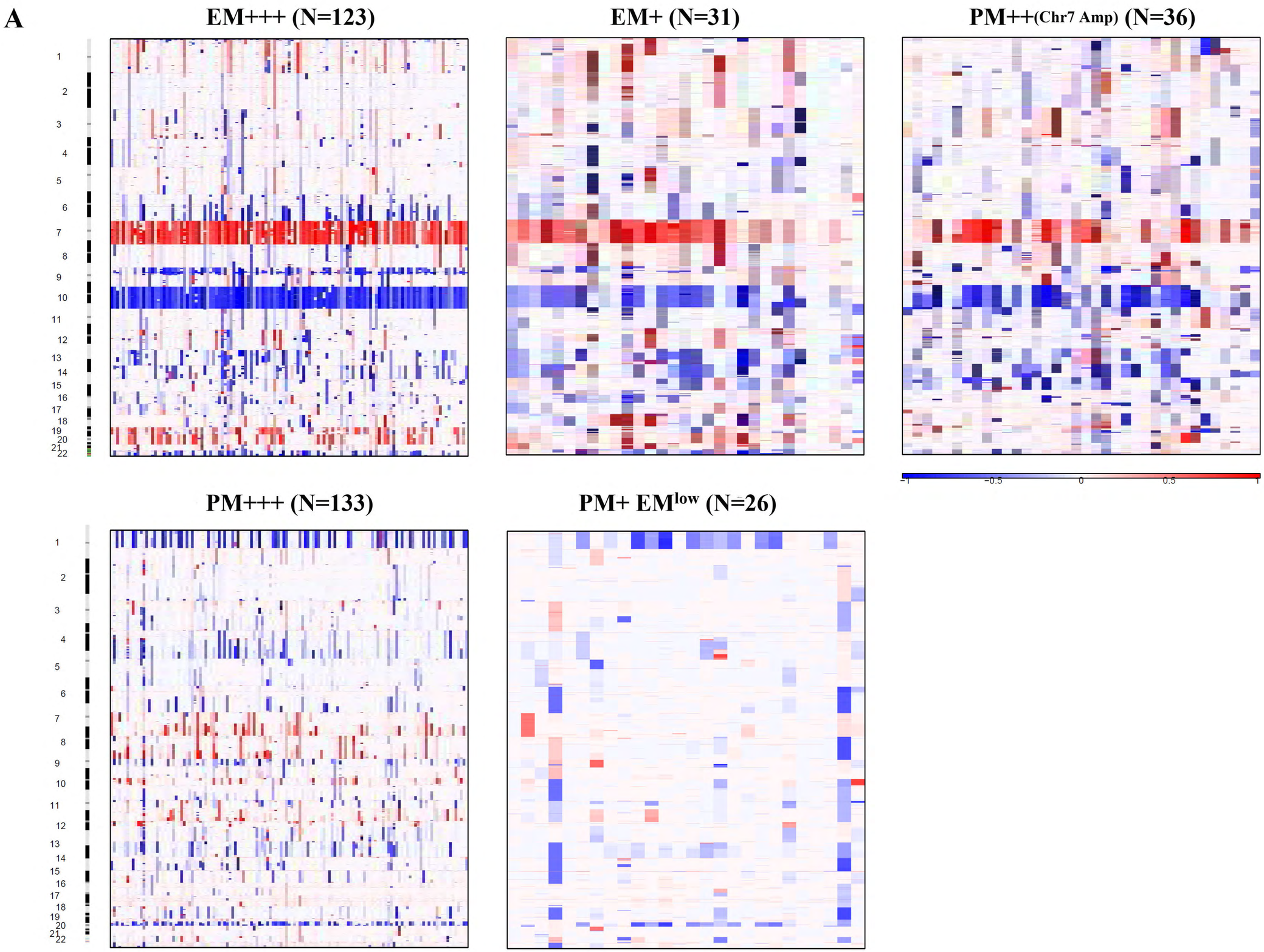

A
EM+++ (N=123)
EM+ (N=31)
PM++(Chr7 Amp) (N=36)
PM+++ (N=133)
PM+ EMlow (N=26)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
-1
-0.5
0
0.5
1

Fig. S11

B

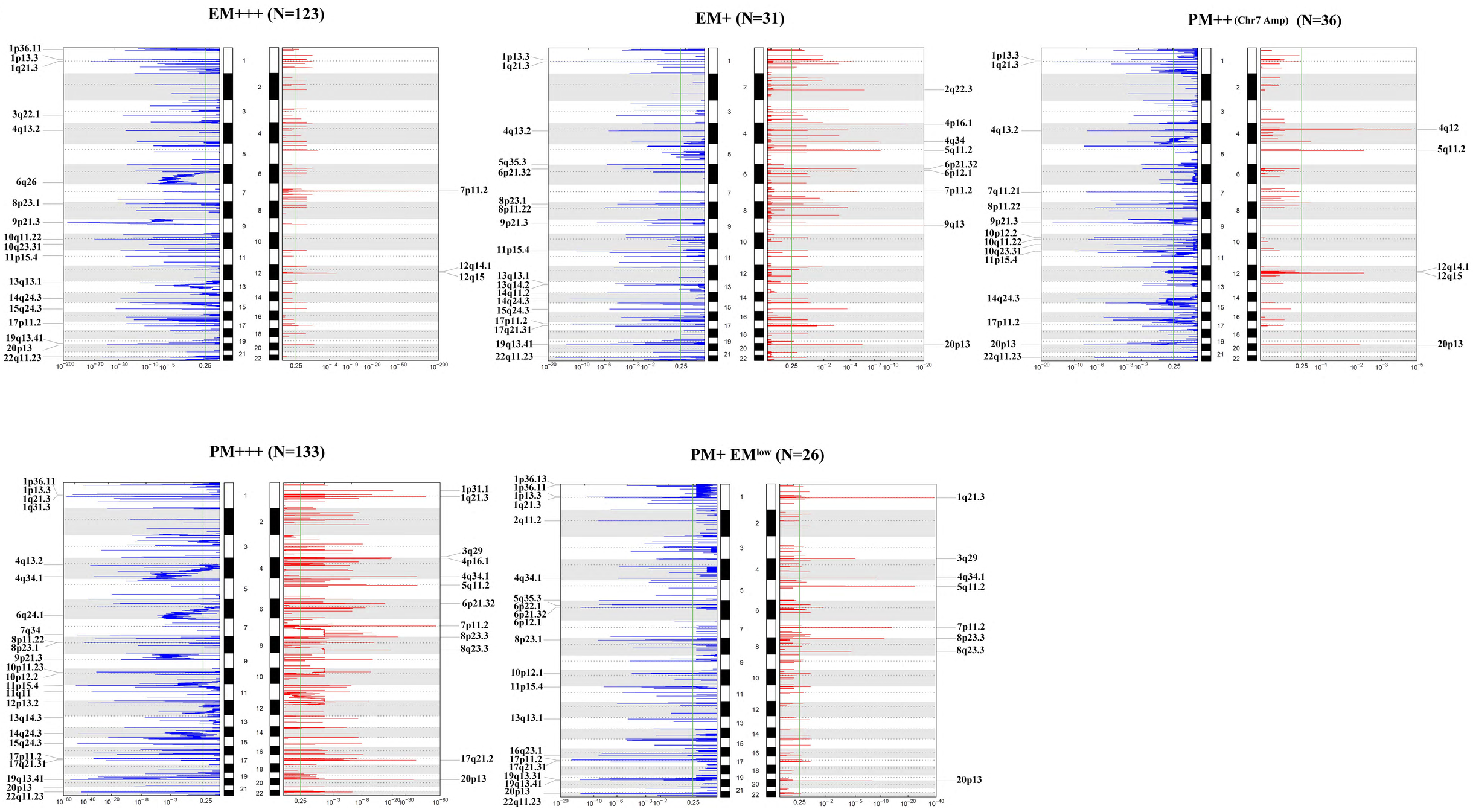

EM+++ (N=123)
1p36.11
1p13.3
1q21.3
3q22.1
4q13.2
6q26
8p23.1
9p21.3
10q11.22
10q23.31
11p15.4
13q13.1
14q24.3
15q24.3
17p11.2
19q13.41
20p13
22q11.23
7p11.2
12q14.1
12q15
EM+ (N=31)
1p13.3
1q21.3
4q13.2
5q35.3
6p21.32
8p23.1
8p11.22
9p21.3
11p15.4
13q13.1
13q14.2
14q11.2
14q24.3
15q24.3
17p11.2
17q21.31
19q13.41
22q11.23
2q22.3
4p16.1
4q34
5q11.2
6p21.32
6p12.1
7p11.2
9q13
20p13
PM++ (Chr7 Amp) (N=36)
1p13.3
1q21.3
4q13.2
7q11.21
8p11.22
9p21.3
10p12.2
10q11.22
10q23.31
11p15.4
14q24.3
17p11.2
20p13
22q11.23
4q12
5q11.2
12q14.1
12q15
20p13
PM+++ (N=133)
1p36.11
1p13.3
1q21.3
1q31.3
4q13.2
4q34.1
6q24.1
7q34
8p11.22
8p23.1
9p21.3
10p11.23
10p12.2
11p15.4
11q11
12p13.2
13q14.3
14q24.3
15q24.3
17p11.2
17q21.31
19q13.41
20p13
22q11.23
1p31.1
1q21.3
3q29
4p16.1
4q34.1
5q11.2
6p21.32
7p11.2
8p23.3
8q23.3
17q21.2
20p13
PM+ EM^low (N=26)
1p36.13
1p36.11
1p13.3
1q21.3
2q11.2
4q34.1
5q35.3
6p22.1
6p21.32
6p12.1
8p23.1
10p12.1
11p15.4
13q13.1
16q23.1
17p11.2
17q21.31
19q13.31
19q13.41
20p13
22q11.23
1q21.3
3q29
4q34.1
5q11.2
7p11.2
8p23.3
8q23.3
20p13

Fig. S12

A

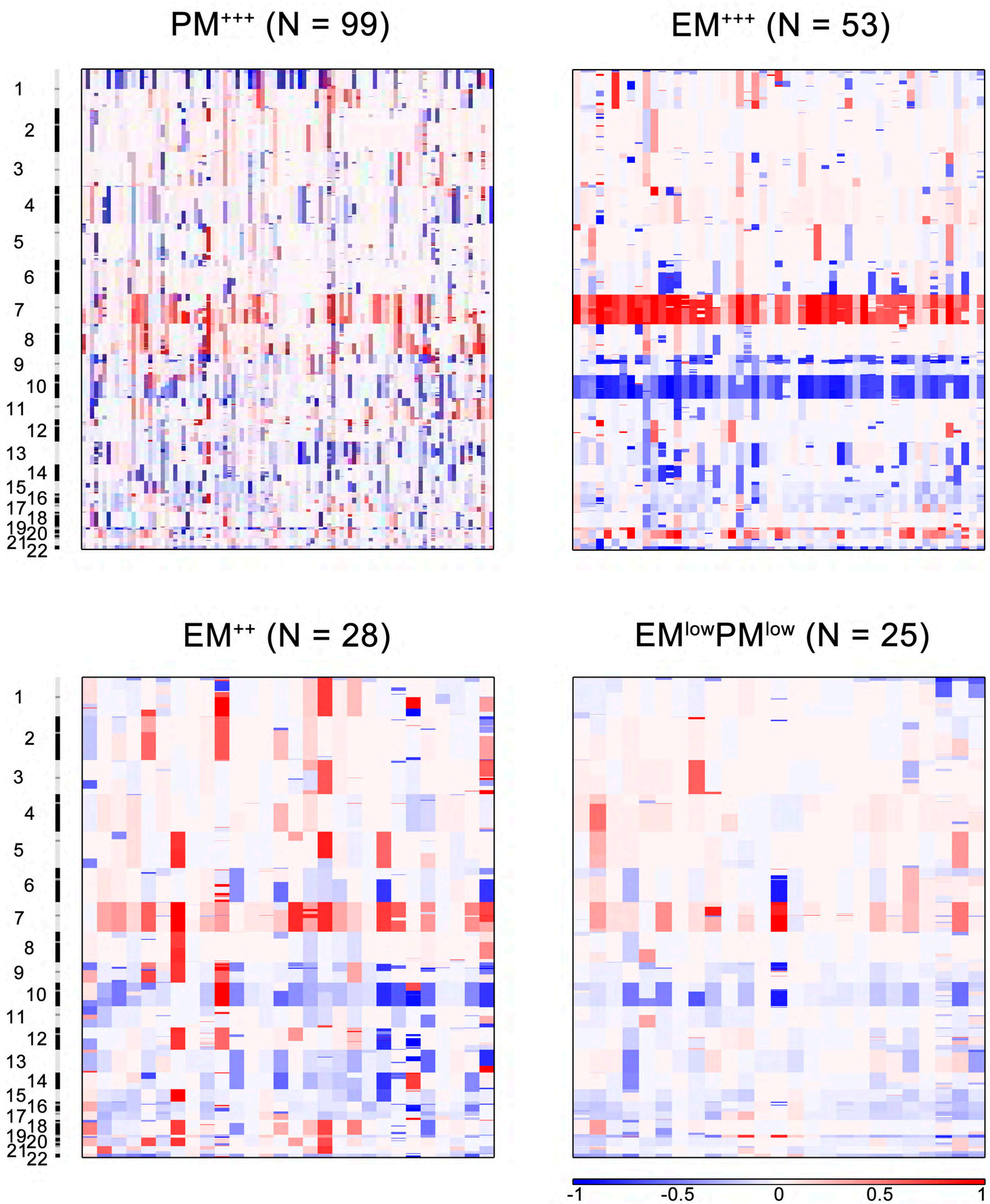

PM+++ (N = 99)
EM+++ (N = 53)
EM++ (N = 28)
EMlowPMlow (N = 25)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
-1
-0.5
0
0.5
1

Fig. S12

B

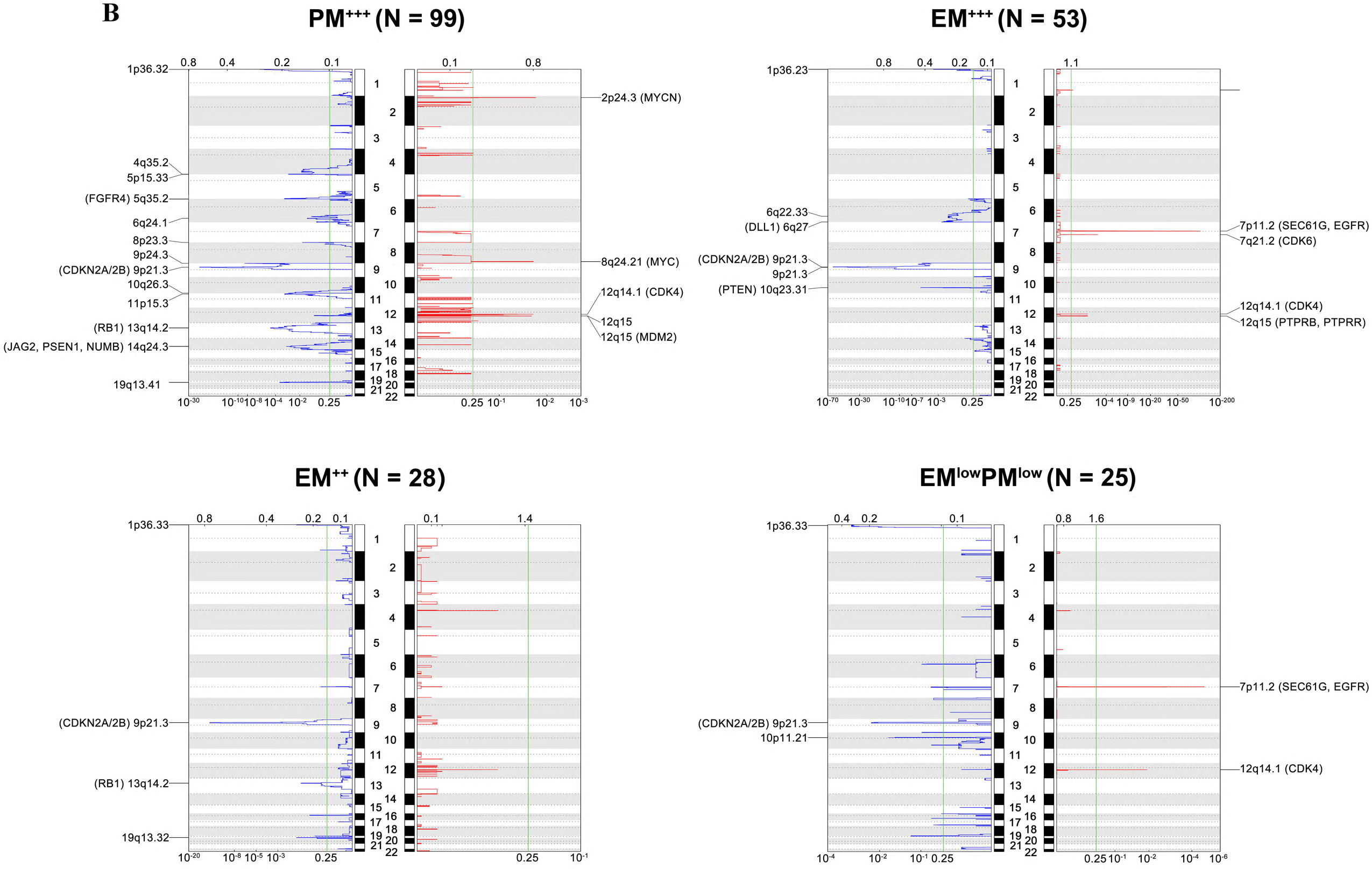
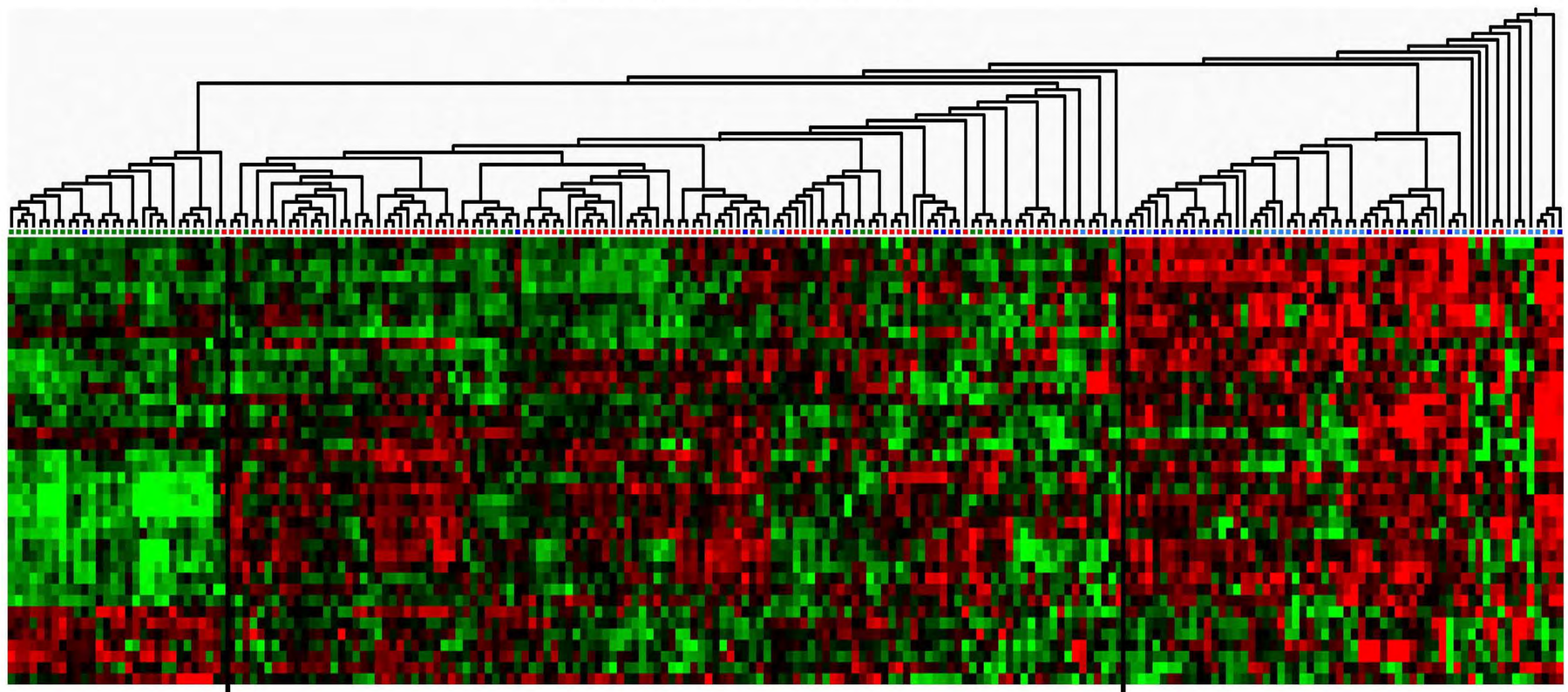

PM+++ (N = 99)
1p36.32
4q35.2
5p15.33
(FGFR4) 5q35.2
6q24.1
8p23.3
9p24.3
(CDKN2A/2B) 9p21.3
10q26.3
11p15.3
(RB1) 13q14.2
(JAG2, PSEN1, NUMB) 14q24.3
19q13.41
2p24.3 (MYCN)
8q24.21 (MYC)
12q14.1 (CDK4)
12q15
12q15 (MDM2)
EM+++ (N = 53)
1p36.23
6q22.33
(DLL1) 6q27
(CDKN2A/2B) 9p21.3
9p21.3
(PTEN) 10q23.31
7p11.2 (SEC61G, EGFR)
7q21.2 (CDK6)
12q14.1 (CDK4)
12q15 (PTPRB, PTPRR)
EM++ (N = 28)
1p36.33
(CDKN2A/2B) 9p21.3
(RB1) 13q14.2
19q13.32
EMlowPMlow (N = 25)
1p36.33
(CDKN2A/2B) 9p21.3
10p11.21
7p11.2 (SEC61G, EGFR)
12q14.1 (CDK4)

Fig. S13

Tiantan Data set

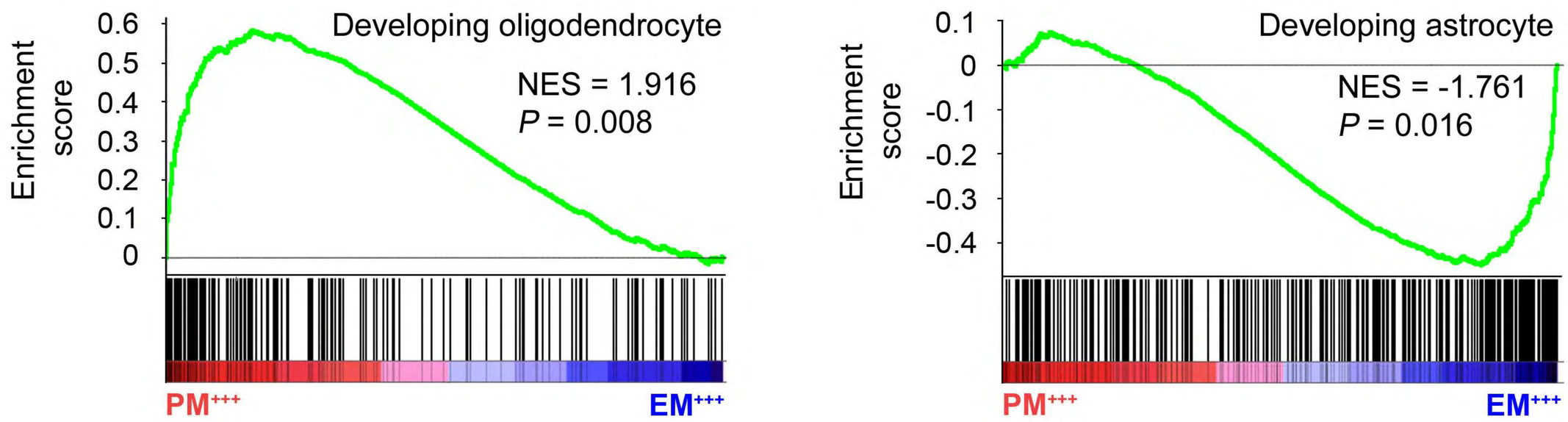
A

M13C (FDR = 0.003)



PM+++	1	92	13
EM+++	1	9	19
EM++	0	5	25
EMlowPMlow	28	16	3

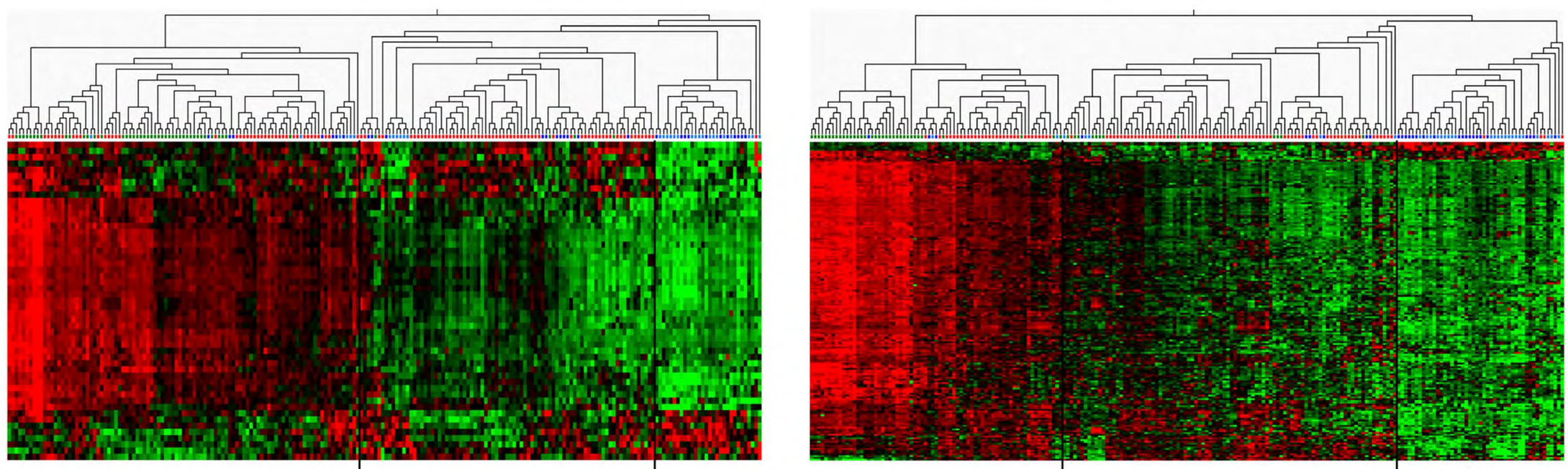
B



C

Mature Oligodendrocyte Signature
(FDR = 1.8×10^{-9})

Mature Neuron Signature
(FDR = 2.8×10^{-13})

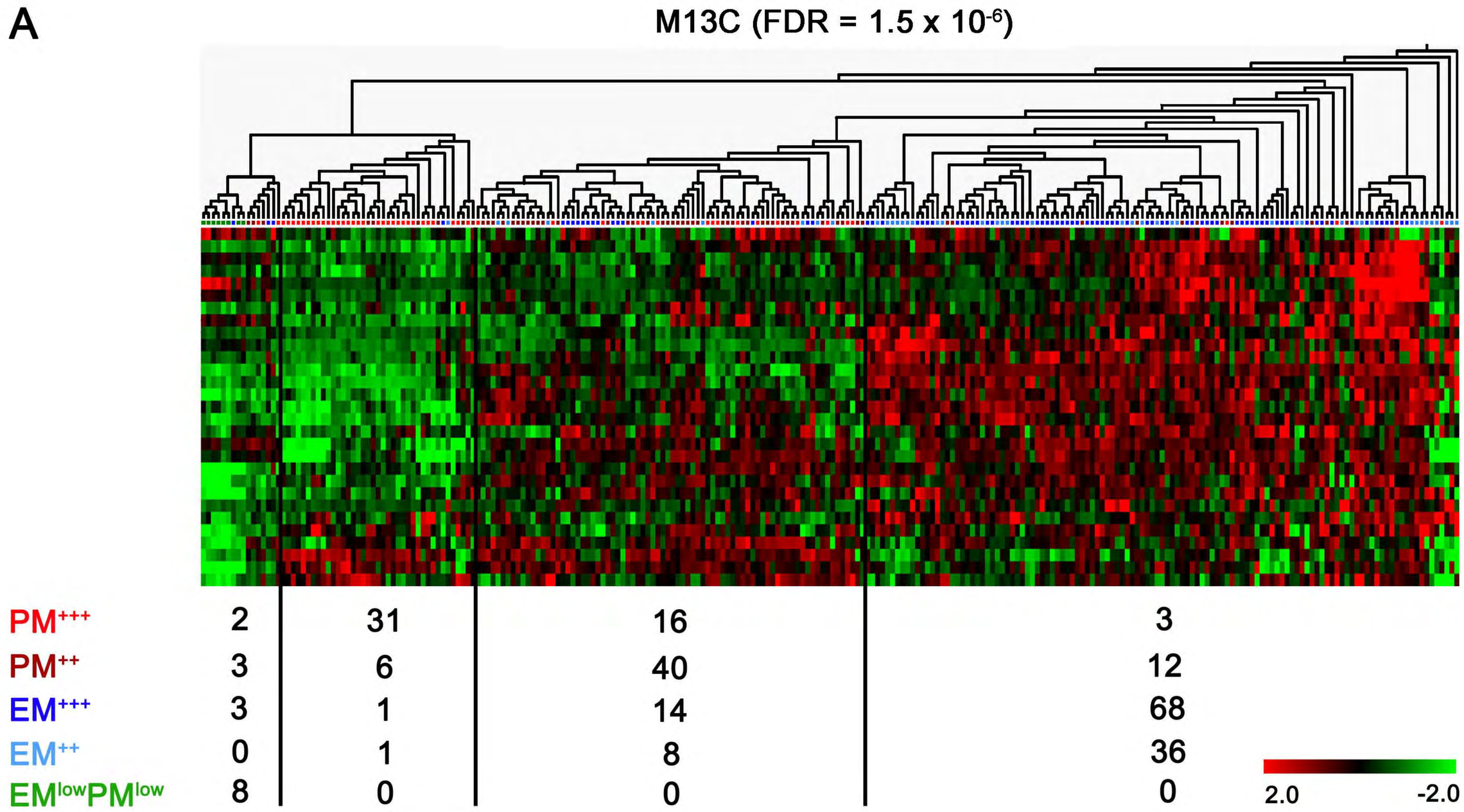


PM+++	43	62	1	31	71	4
EM+++	7	11	11	3	6	20
EM++	6	6	18	2	5	23
EMlowPMlow	43	4	0	35	12	0

Fig. S13

GSE16011 Data set

A



B

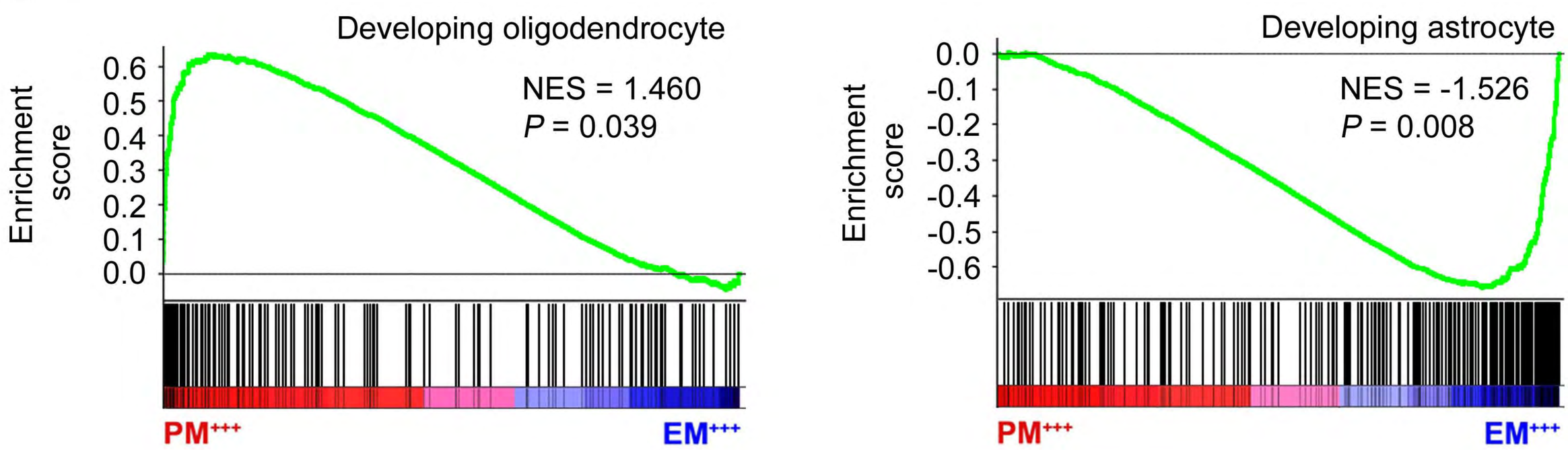
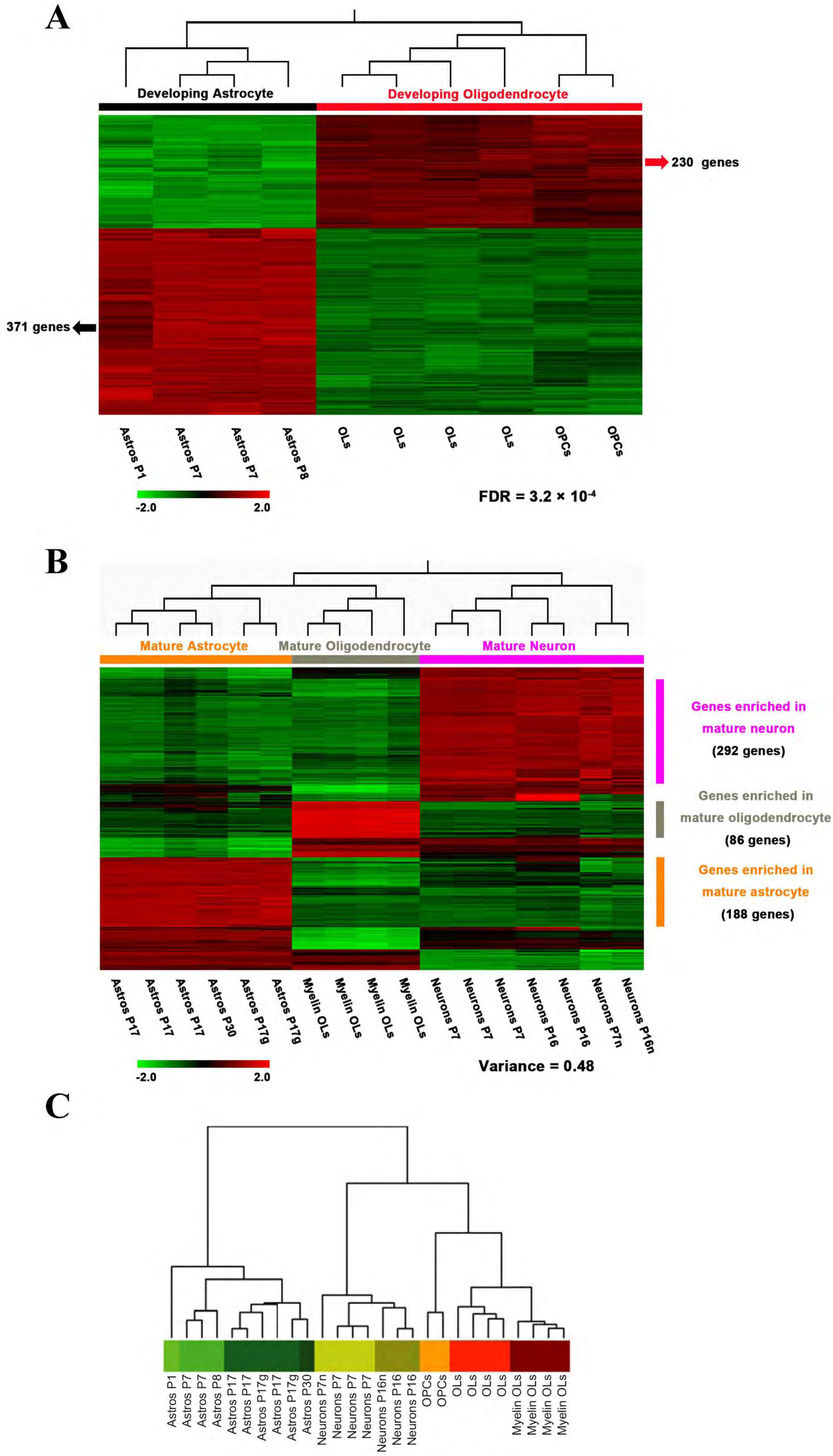


Fig. S14



Cohoy et al. J Neurosci 28:264,2008.

Table S1. Glioma data sets analyzed in this study.

Data set	Platform	Composition
1) Tiantan*	Agilent 44K array	212 glioma (AII 58, AIII 8, GBM 82, OII 18, OIII 11, OAII 20, OAIII 15)
2) GSE4290[1]	HG-U133 Plus 2.0	157 glioma (AII 7, AIII 19, GBM 77, OII 38, OIII 12, unknown 4), 23 epilepsy
3) GSE16011[2]	HG-U133 Plus 2.0	247 glioma (PA 3, AII 12, AIII 16, GBM 142, OII 8, OIII 39, OAII 3, OAIII 24), 8 normal adult brain samples
4) REMBRANDT[3]**	HG-U133 Plus 2.0	521 gliomas (A 148, GBM 228, O 67, OA 11, unknown 67), 21 epilepsy
5) TCGA mRNA-seq	IlluminaHiSeq_RNASeq	381 gliomas (AII 17, AIII 45, GBM 161, OII 25, OIII 29, OAII 52, OAIII 37, unknown 15)

The abbreviations for tumor types were derived from the source data: A, astrocytoma, AII, astrocytoma grade II; AIII, astrocytoma grade III; GBM, glioblastoma; O, oligodendroglioma; OII, oligodendroglioma grade II; OIII, oligodendroglioma grade III; OA, oligoastrocytoma; OA II, oligoastrocytoma grade II; OA III, oligoastrocytoma grade III; PA, pilocytic astrocytoma.

*: Newly established at Tiantan Hospital in Beijing, the study was approved by the Ethics Committee of Beijing Tiantan Hospital and written informed consent was obtained from all patients. The gene expression data, and the clinical and pathological parameters of all gliomas in the Tiantan data set can be retrieved from the Chinese Glioma Genome Atlas (http://www.cgga.org.cn). **: the gene expression data of the REMBRANDT data set were downloaded in April 2011 and its clinical data in May 2012. The data for the TCGA mRNA-seq data set was downloaded in April 2013.

1. Sun, L. *et al.* Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell* **9**, 287-300 (2006).
2. Gravendeel, L.A. *et al.* Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res* **69**, 9065-72 (2009).
3. Madhavan, S. *et al.* Rembrandt: helping personalized medicine become a reality through integrative translational research. *Mol Cancer Res* **7**, 157-67 (2009)

Table S2. Members of EM and PM classifiers.

EM

Regulation of nervous system development: NES, SHOX2, SEMA6D, SALL1, EGFR

Regulation of astrocyte development: NFIA, SOX9, POU3F2, SHOX2, CDKN2C, UHRF1, EGFR

JAK-STAT cascade/regulation of growth hormone signal transduction: SOCS2

JNK cascade: TNFRSF19

Regulation of apoptosis: SOCS2, TNFRSF19

Regulation of WNT receptor signaling pathway: SALL1

Regulation of TGF-beta signaling: TGIF2

Regulation of NOTCH signaling: LFNG

Metabolic process: ACSS3, ELOVL2, SPPL2B, PRPF31, LFNG

Regulation of mRNA processing or transcription: PRPF31, ZNF320, ZNF558, ZNF45, ZNF551, DMRTA2, POU3F2, UHRF1

Cell adhesion and migration, cytoskeleton organization and protein transport: ITGB8, TRIOBP, VAV3, SNX5, DENND2A, SEC61G

Extracellular matrix heparan sulfate proteoglycan metabolism: HS3ST3B1

Cell cycle/cell proliferation/DNA repair: CDKN2C, UHRF1, POU3F2, PDGFA

Angiogenesis: VAV3

Proteasomal ubiquitn-dependent protein catabolic process: RNF180

Genes of unknown function: LOC144481, NCOA3, PMS2P11, MGC12916

9 unknown genes

PM

Pattern formation/multicellular organismal development: TSHZ1, OLIG1, PCGF2

Nervous system development: MYT1, NAV1, FAM5C, CHD7, ZEB1, NLGN1, TACC2, LPPR1/RP11-35N6.1

Oligodendroyte cell fate specification and differentiation: SOX6, SOX8, OLIG2, MYT1, PDGFRA

Neuronal fate commitment: OLIG2, OLIG1

Regulation of axon genesis: VAX2

mRNA processing: FIP1L1

Canonical Wnt receptor signaling pathway: SOX4

Regulation of transcription and mRNA processing: FIP1L1, ZNF462, ZFP37, ETV1, ZNF22, EIF4EBP2, NOVA1, MEX3A, POLR2F, PATZ1, RFXDC2/RFX7

Collagen catabolic process: MMP16

Calcium ion binding: C9orf140, CACNG4

DNA repair/cell proliferation: CSNK1E, TACC2

Cell adhesion and migration, cytoskeleton organization: MARCKS, NLGN1, TACC2, NOVA1

Cellular defense response: KLRC3

Ubiquitin-dependent protein catabolic process: CSNK1E

G-protein coupled receptor protein signaling pathway///neuropeptide signaling pathway: LPHN3

Locomotory behavior: C1QL1

Genes of unclear functions: LIX1L, NXPH1, LOC283174, LOC100507187, C10orf18, C1orf106, TEX10

4 unknown genes

Members of EM and PM were identified with decreasing correlation index in GSE4290 using Qlucore Omics Explorer 2.2. The probe sets 201983_s_at (EGFR) and 203131_at (PDGFRA) were used to search for the top 68 and 56 most correlated probe sets containing 37 and 44 correlated known genes in EM and PM, respectively. Genes are categorized according to GO processes.

Table S3. Clinical and genomic characteristics of EM/PM glioma subtypes.

Data set		Subtypes					P value
Tiantan		**$EM^{low}PM^{low}$**	**PM^{+++}**	**EM^{+++}**	**EM^{++}**		
(N=212)	Number of samples	47	106	29	30		
	GBM	10	26	**19**	**25**		< 0.0001
	Astrocytoma III	2	3	3	0		
	Astrocytoma II	16	**39**	1	2		0.0001
	Oligoastrocytoma III	5	6	4	0		
	Oligoastrocytoma II	5	15	1	0		
	Oligodendroglioma III	2	6	0	2		
	Oligodendroglioma II	6	11	0	0		
	IDH1 mutation	16/39	**67/91**[a]	0/20	2/27		
	Age at diagnosis (years; median ± SD)	39.0± 12.2	38.0 ± 10.7	50.0± 9.3	40.5± 14.6		0.0004
	Survival years (median± SD)	2.0± 0.9	2.3± 0.8	1.2± 0.7	1.1± 0.6		< 0.0001
GSE16011		**$EM^{low}PM^{low}$**	**PM^{+++}**	**PM^{++}**	**EM^{+++}**	**EM^{++}**	
(N=252)	Number of samples	8	52	61	86	45	
	GBM	0	10	26	**70**	**36**	< 0.0001[b]
	Astrocytoma III	0	1	8	3	4	
	Astrocytoma II	0	2	9	0	1	
	Oligoastrocytoma III	0	7	6	10	1	
	Oligoastrocytoma II	0	1	2	0	0	
	Oligodendroglioma III	0	**27**	6	3	3	< 0.0001[b]
	Oligodendroglioma II	0	4	4	0	0	
	Non-tumor	8	0	0	0	0	
	1p LOH	0/4	**32/45**[a]	10/29	7/42	3/18	
	19q LOH	0/4	**32/45**[a]	7/29	6/38	4/18	
	IDH1 mutation	2/5	**27/41**[a]	24/48	11/68	11/37	
	EGFR amplification	0/6	0/36	2/29	**30/47**[a]	5/21	

Primary GBM	0	18	32	**59**	27	0.001 [b]
Secondary GBM	0	**31**	25	24	16	0.003 [b]
Age at diagnosis (years; median ± SD)		45.1 ± 12.6	43.0 ± 13.7	56.9 ± 12.2	48.0 ± 14.8	< 0.0001
Survival years (median± SD)		3.9 ± 4.2	2.0 ± 4.3	0.7 ± 0.6	1.2 ± 2.4	< 0.0001

REMBRANDT		$EM^{low}PM^{low}$	PM^{+++}	EM^{+++}	EM^{++}	
(N=542)	Number of samples	86	230	149	77	
	GBM	34	48	**102**	**44**	< 0.0001
	Astrocytoma	15	**99**	23	11	< 0.0001
	Oligoastrocytoma	1	7	0	3	
	Oligodendroglioma	8	**47**	10	2	< 0.0001
	Unknown	7	29	14	17	
	Non-tumor	21	0	0	0	
	Amp 7p11/Del 10q23	0	0	46/53	0/28	
	1p36/19q13 co-deletion	0	44/99	0/53	4/28	
	Survival years (median± SD)	1.8 ± 2.0	3.2 ± 3.9	1.0 ± 1.2	1.3 ± 1.7	< 0.0001

Age at diagnosis and survival period were analyzed using one-way ANOVA and log-rank (Mantel-Cox) tests, respectively. Distribution of morphological diagnosis among EM/PM glioma subtypes was analyzed using the Pearson χ^2 test. **a**: $p < 0.0001$, frequencies of characteristic genomic alterations in the indicated group compared with the other groups in combination using two-sided Fisher's exact test. **b**: Frequencies of primary and secondary GBM in all subtypes except $EM^{low}PM^{low}$ were analyzed using the Pearson χ^2 test. Genomic alterations of amplification at chromosome 7p11 coupled with deletion at chromosome 10q23 include amplification of *EGFR* and deletion of *PTEN* in glioma genomes. Gliomas in the REMBRANDT data set are presented without grade information due to a lack of grade information in 109

samples.

Table S4. Genes enriched in EM/PM glioma subtypes.

Genes enriched in EM gliomas	
GO:Process	**Gene symbol**
regulation of cell development and maturation	MEOX2, TWSG1, ANXA2, SHOX2, TIMP1, TGIF1, TAGLN2, PDPN, FNDC3B, CAST
cell-cell adhesion, cell motility	CD58, LAMC1, MSN, CMTM3, TRIP6
apoptosis-related	ANXA1, ANXA5, STEAP3, CASP6
intracellular transport of mRNA or protein	NUP37, KDELR2, SEC61G, RAB34
carbohydrate metabolic process	CHI3L1, GUSB
cell growth	EMP3, IGFBP2
G-protein coupled receptor protein signaling pathway	GNAI3, F2R, FZD1
metabolic process	ACSS3, METTL7B
positive regulation of I-kappaB kinase/NF-kappaB cascade	TSPAN6
protein folding	PPIC
regulation of small GTPase mediated signal transduction	IQGAP1
regulation of transcription, DNA-dependent	ZNF217, PTRF
response to stress///response to unfolded protein	SERPINH1, CLIC1
response to virus	ZC3HAV1
RNA splicing	PTBP1
blood coagulation	EFEMP2
BMP signaling pathway	FSTL1
cell redox homeostasis///protein secretion	PDIA4
apical protein localization///cell morphogenesis	SHROOM3
cytoskeleton organization and biogenesis	FLNA
uncharacterized processes	ANTXR2, C7orf42, CCDC109B, DPY19L1, SPPL2A, DYNLT3, FAM129A

Genes enriched in PM gliomas

GO:Process	Gene symbol
organ development	CHD7
protein O-linked glycosylation	GALNT13
BMP signaling pathway, positive regulation of astrocyte differentiation	BMP2
cell adhesion///nervous system development	DSCAM
cellular defense response	KLRC3
chromatin modification	H2AFY2, CHD7
ion transport///sodium ion transport	SCN3A, FXYD6
locomotory behavior///phosphate transport	C1QL1
metabolic process	GSTA4
microtubule-based movement	KIF21B
neuron fate commitment///oligodendrocyte differentiation	OLIG2
negative regulation of Wnt receptor signaling pathway	CXXC4
Notch signaling pathway	DLL1, DLL3, ASCL1, MAML2
regulation of transcription, DNA-dependent///transcription	SOX4, SOX8, SOX6, MYT1, BCL7A, TCF12, OLIG2
uncharacterized processes	FAM110B, FCHSD2, MEGF11, SHD, SMOC1

Genes enriched in $EM^{low}PM^{low}$ gliomas

GO:Process	Gene symbol
Cytoskeleton organization, cell adhesion, Signal transduction	TNXB, TRIP10, TTLL3, SYDE1, ITGB4, ROM1, RHOQ, DOK1, RGS1, ITSN1, MAP4, DTNA, NOTCH2NL
Metabolic process	ALDH3B1, LPL, HYAL2, LASS1, PLTP, PRCP, CAPN2
DNA replication, regulation of transcription, cell cycle, apoptosis	TNFRSF14, NFKBIA, DDIT4, HMG20B, NUPR1, NFIC, PER1, RASSF4, IGFBP7, ZHX3, TSC22D3, CEBPD
Translation	RPS27L
Regulation of development, cell growth	PLOD1, AGXT2L2, METRN, FAM107A
Protein folding, protein transport	SCAMP2, FKBP5, SCARNA16
Response to stress	TMEM204

Uncharacterized processes	PRRG4, GLTSCR2, FAM14A, TNRC6A, MSI2, GPR108, NBPF12, TMEM150, LYAR, C3orf15, AFF1, WDR27, SHARPIN

Genes enriched in EM and PM gliomas were selected by their common enrichment in GSE16011, Tiantan and the REMBRANDT data sets as in the analyses shown in Fig. S4A. Genes enriched in $EM^{low}PM^{low}$ gliomas were selected by their differential expression between the $EM^{low}PM^{low}$ gliomas and the control brain tissues in the REMBRANDT data sets as depicted in Fig. S4B.

Table S5. Hazard ratios for association between EM/PM scores and the overall mortality using multivariate Cox regression analysis

Data set	Variate	P value	Hazard ratio	95% CI*
Tiantan	EM score	0.001	2.28	1.40-3.69
	PM score	0.135	0.71	0.46-1.11
	IDH1 mut.	0.139	0.60	0.30-1.18
	Age	0.917	1.00	0.98-1.02
	Gender	0.473	0.84	0.52-1.36
	KPS	2.207×10^{-11}	0.95	0.94-0.96
GSE16011	EM score	0.053	1.76	0.93-3.11
	PM score	0.026	0.60	0.38-0.94
	IDH1 mut.	0.626	0.86	0.48-1.54
	1p19q del.	0.068	0.60	0.35-1.03
	Age	0.068	1.02	0.99-1.03
	Gender	0.189	0.71	0.42-1.18
	KPS	5.086×10^{-5}	0.97	0.94-0.98
Rembrandt	EM score	0.050	1.59	1.00-2.54
	PM score	0.011	0.49	0.29-0.85
	1p19q del.	0.078	0.49	0.22-1.08
	Age	0.009	1.14	1.03-1.27
	Gender	0.600	0.85	0.48-1.52
TCGA mRNA-seq	EM score	0.075	1.25	0.98-1.60
	PM score	0.352	0.89	0.71-1.13
	IDH1 mut.	4.303×10^{-4}	0.32	0.17-0.60
	1p19q del.	0.080	0.57	0.31-1.06
	Age	2.382×10^{-10}	1.05	1.04-1.06
	Gender	0.998	1.00	0.70-1.43

EM and PM scores for glioma samples were calculated as the average of normalized expression levels for all EM or PM signature genes detected in each data set. For the expression levels of individual EM/PM members, probe sets not concordant with EM/PM clustering were excluded. Multivariate Cox regression analysis was used to calculate the hazard ratio and 95% confidence interval (CI), adjusted for other confounding factors, including age (continuous, except in the REMBRANDT data set), gender (male and female), IDH1 mutation (yes and no), 1p19q deletion (yes and no), and KPS (continuous). All the analyses were performed using SPSS software. Age was converted to 5-year age group (categorical variable) for the analysis of the REMBRANDT data set. KPS data was not available for many patients in the REMBRANDT and TCGA mRNA-seq data sets and therefore not included in the analysis of these two data sets.

Table S6. Cross-labeling of glioma samples in NMF clustering by TCGA or EM/PM classifiers.

TCGA cluster	EM/PM clusters				
Tiantan					
	$EM^{low}PM^{low}$	**PM^{+++}**		**EM^{+++}**	**EM^{++}**
PN	0	67		0	1
NL	40	6		3	0
CL	3	27		5	5
Mes	4	6		21	3
GSE16011					
	$EM^{low}PM^{low}$	**PM^{+++}**	**PM^{++}**	**EM^{+++}**	**EM^{++}**
PN	8	11	25	3	0
NL	0	40	25	1	6
CL	0	0	1	76	15
Mes	0	1	10	6	24
REMBRANDT					
	$EM^{low}PM^{low}$	**PM^{+++}**		**EM^{+++}**	**EM^{++}**
PN	0	173		0	2
NL	85	38		10	9
CL	0	1		92	4
Mes	1	18		47	62

Using NMF, the Tiantan, GSE16011 and REMBRANDT data sets were re-classified according to the TCGA classifiers. *PN* and *NL* gliomas were associated with better survival compared to *Mes* and *CL* gliomas. *PN* and *CL* gliomas predominantly contained PM and EM^{+++} gliomas, respectively. Previous reports (Int. J. Cancer 122:178-1786, 2008; Molecular Cancer 7:41, 2008; J Clin Oncol 24:4758-4763, 2006; Cancer Res 65:1678-1686, 2005; Neurology 74:1886-1990, 2010; J Clin Oncol 29:4482-4490, 2011) and our EM/PM clustering demonstrated that gliomas with EGFR amplification/PTEN deletion (in >90% of EM^{+++} gliomas) or with co-deletion of 1p36/19q13 (in ~40% of PM^{+++} gliomas) represent mutually exclusive subtypes regarding genomic, transcriptomal and clinical features. These two types of gliomas were however found in *NL* and *Mes* subtypes.

Table S7. EM/PM gliom-specific gain or loss of EM/PM classifiers in the TCGA mRNA-seq data set

		EM+++ gliomas (n =123)							
		SCNA (n)				Gene dosage effect		modulating capability	
EGFR-M	**Chr**	**Gain**		**Loss**		**R**	**p value**	**β**	**p value**
		all data	**focal data**	**all data**	**focal data**				
EGFR	7p11.2	**119**	**83**	0	2	0.713	<0.0001	0.052	<0.0001
SEC61G	7p11.2	**115**	**68**	0	3	0.849	<0.0001	0.036	<0.0001
ITGB8	7p21.1	**103**	**6**	0	4	0.375	<0.0001	0.115	0.004
LFNG	7p22.3	**98**	**5**	0	7	0.379	<0.0001	0.313	<0.0001
PDGFA	7p22.3	**99**	**7**	1	8	0.304	0.0006	0.484	<0.0001
PMS2P11	7q11.23								
DENND2A	7q34	**103**	2	0	2	0.441	<0.0001	0.381	<0.0001
VAV3	1p13.3	15	1	1	1	No Effect		1.171	0.002
DMRTA2	1p33	15	2	4	4	No Effect		2.347	<0.0001
CDKN2C	1p33	15	2	11	11	0.383	<0.0001	0.355	0.006
NFIA	1p31.3	15	2	1	1	No Effect		No Effect	
ZNF558	19p13.2	**49**	3	2	2	0.704	<0.0001	2.727	<0.0001
UHRF1	19p13.3	**50**	6	1	2	No Effect		1.326	<0.0001
PRPF31	19q13.42	**43**	2	6	4	0.603	<0.0001	1.227	<0.0001
ACSS3	12q21.31	8		11				0.649	0.021
SOCS2	12q22	7		8				0.441	<0.0001
HS3ST3B1	17p12	4		17					
SOX9	17q24.3	7		9				0.271	<0.0001
NES	1q23.1	18		0				0.062	<0.0001
SHOX2	3q25.32	8		7				1.77	0.004
RNF180	5q12.3	6		4				0.877	<0.0001
ELOVL2	6p24.2	2		18				0.345	<0.0001
POU3F2	6q16	3		34				0.454	0.003
TNFRSF19	13q12.12	3		33				0.825	<0.0001
SEMA6D	15q21.1	2		19				0.403	0.001
SALL1	16q12.1	1		14				0.66	<0.0001

SNX5	20p11	**42**	0	1	1	0.556	<0.0001	No Effect	
NCOA3	20q12	**43**	1	0	0	0.33	0.0002	1.488	0.013
TRIOBP	22q13.1	3	3	49	8	0.537	<0.0001	1.541	0.002
PDGFRA-M									
PDGFRA	4q12	13		6					
LPHN3	4q13.1	1		8					
-35N6.1 (LP	9q31.1	9		17					
ZNF462	9q31.2	10		16					
rf140 (SAPC	9q34.3								
LIX1L	1q21.1	16		1					
MEX3A	1q22	18		0					
FAM5C	1q31.1	17		2					
C1orf106	1q32.1	15		3					
NAV1	1q32.1	15		3					
OLIG1	21q22.11	9		5					
OLIG2	21q22.11	9		5					
PATZ1	22q12.2	2	1	**47**	7	0.288	0.0012	0.906	<0.0001
CSNK1E	22q13.1	2	1	**50**	8	0.485	<0.0001	0.408	<0.0001
POLR2F	22q13.1	1	1	**49**	8	0.418	<0.0001	No Effect	
PCGF2	17q12	3		16					
C1QL1	17q21.31	5		10					
CACNG4	17q24.2	7		6					
TMCC1	3q22.1	9		7					
NLGN1	3q26.31	13		7					
SOX4	6p22.3	4		13					
MARCKS	6q21	1		36					
ETV1	7p21.2	**102**	8	0	4	0.298	0.0008	0.203	<0.0001
NXPH1	7p21.3	**100**	5	1	5	No Effect		0.719	<0.0001
CHD7	8q12.1	5		8					
MMP16	8q21.3	5		6					

orf18 (FAM2(	10p15.1	23		27					
ZEB1	10p11.22	1	2	**118**	0	-0.32	0.0003	0.365	<0.0001
ZNF22	10q11.21	0	2	**119**	0	-0.407	<0.0001	1.047	<0.0001
EIF4EBP2	10q22.1	1	1	**119**	2	No Effect		No Effect	
TACC2	10q26.13	2	0	**121**	2	-0.382	<0.0001	1.064	<0.0001
SOX6	11p15.2	4		10					
LOC283174	11q25	2		12					
KLRC3	12p13.2	9		10					
NOVA1	14q12	1	1	**36**	10	0.417	<0.0001	0.481	<0.0001
RFX7	15q21.3	3		16					
SOX8	16p13.3	1		10					
TSHZ1	18q22.3	5		8					
MYT1	20q13.33	42	1	0	3	No Effect		1.268	<0.0001

PM+++ gliomas (n =133)

		SCNA (n)				Gene dosage effect		mudulated capability	
EGFR-M	**Chr**	**Gain grade (n)**		**Loss grade (n)**					
		all data	**focal data**	**all data**	**focal data**	**R**	**p value**	**B**	**p value**
EGFR	7p11.2	16	3	0	3	No Effect		0.037	0.002
SEC61G	7p11.2	16	3	0	3	0.222	0.0104	No Effect	
ITGB8	7p21.1	12	2	1	5	No Effect		0.118	<0.0001
LFNG	7p22.3	12	2	6	9	No Effect		0.211	<0.0001
PDGFA	7p22.3	12	2	6	9	No Effect		No Effect	
PMS2P11	7q11.23								
DENND2A	7q34	34	14	0	0	0.245	0.0046	0.421	<0.0001
VAV3	1p13.3	2	1	**53**	4	No Effect		No Effect	
DMRTA2	1p33	1		**52**	3	No Effect		No Effect	
CDKN2C	1p33	8		**30**	5	No Effect		0.216	0.004
NFIA	1p31.3	1		**53**	4	0.7	<0.0001	0.189	<0.0001
ZNF558	19p13.2	7		1		0.3376	<0.0001	0.945	<0.0001

UHRF1	19p13.3	6		2		No Effect		0.424	<0.0001
PRPF31	19q13.42	0		**79**	**20**	0.809	<0.0001	0.48	<0.0001
ACSS3	12q21.31	1		15					
SOCS2	12q22	1		14					
HS3ST3B1	17p12	2		0					
SOX9	17q24.3	5		0				0.108	<0.0001
NES	1q23.1	8		0				0.047	<0.0001
SHOX2	3q25.32	1		11					
RNF180	5q12.3	0		3				0.479	<0.0001
ELOVL2	6p24.2	4		5					
POU3F2	6q16	1		12				0.582	<0.0001
TNFRSF19	13q12.12	0		25				0.438	<0.0001
SEMA6D	15q21.1	1		9					
SALL1	16q12.1	1		3				0.336	<0.0001
SNX5	20p11	4		2		No Effect		0.364	<0.0001
NCOA3	20q12	2		0		No Effect		0.537	<0.0001
TRIOBP	22q13.1	2		7		No Effect		0.809	<0.0001

PDGFRA-M

PDGFRA	4q12	5	5	15	1
LPHN3	4q13.1	4	4	19	3
-35N6.1 (LP	9q31.1	3	1	15	1
ZNF462	9q31.2	4		6	
rf140 (SAPC	9q34.3	13		8	
LIX1L	1q21.1	5		2	
MEX3A	1q22	8		0	
FAM5C	1q31.1	3		3	
C1orf106	1q32.1	3		4	
NAV1	1q32.1	2		4	
OLIG1	21q22.11	5		4	
OLIG2	21q22.11	5		4	

PATZ1	22q12.2	2		5		No Effect	1.652	<0.0001
CSNK1E	22q13.1	2		7		No Effect	0.399	<0.0001
POLR2F	22q13.1	2		7		No Effect	0.619	<0.0001
PCGF2	17q12	2		0				
C1QL1	17q21.31	2		0				
CACNG4	17q24.2	4		0				
TMCC1	3q22.1	7		5				
NLGN1	3q26.31	3		7				
SOX4	6p22.3	3		3				
MARCKS	6q21	1		17				
ETV1	7p21.2	14	3	1	6	No Effect	0.121	<0.0001
NXPH1	7p21.3	15	3	1	5	No Effect	0.463	<0.0001
CHD7	8q12.1	16		0				
MMP16	8q21.3	19		0				
orf18 (FAM2(	10p15.1	22		33				
ZEB1	10p11.22	21	12	2		0.0062	0.304	<0.0001
ZNF22	10q11.21	0		2		No Effect	0.312	<0.0001
EIF4EBP2	10q22.1	1		4		0.044	0.461	<0.0001
TACC2	10q26.13	0		17	14	0.0001	1.586	<0.0001
SOX6	11p15.2	7		16				
LOC283174	11q25	23		3				
KLRC3	12p13.2	15		1				
NOVA1	14q12	3		4		0.0004	1.009	<0.0001
RFX7	15q21.3	1		11				
SOX8	16p13.3	2		2				
TSHZ1	18q22.3	1		14				
MYT1	20q13.33	2		3		0.0006	0.851	<0.0001

The segmented copy number data of 600 SNP array data from the TCGA mRNA-seq cohort were analyzed with GISTIC 2.0 at a threshold of 0.2.

N. T.: alteration not detected.

Concomitant gains or losses of genes are labelled with color. Bold labelled genes are enriched in developing astrocytes or OPCs as analyzed in Fig 3.

Gene dosage-dependent expression of EM/PM classifiers was analyzed using Spearman´s rank correlation, correlation coefficients (R) and their p values are indicated.

The modulation capacity of individual EM/PM classifiers to the EM/PM signature was calculated using linear regression analyses between the expression of individual classifiers and EM/PM signature in glioma samples. Regression coefficients (β) and their p values are indicated.

Table S8. Recurrent and concomitant gain or loss of EM/PM claasifier genes in glioma genomes as analyzed in the REMBRANDT data set

		EM+++ (n = 53)		EM++ (n = 28)		PM+++ (n =99)		EMlowPMlow (n = 25)	
EGFR-M	**Chr**	**Gain grade (n)**	**Loss grade (n)**	**Gain grade (n)**	**Loss grade (n)**	**Gain grade (n)**	**Loss grade (n)**	**Gain grade (n)**	**Loss grade (n)**
EGFR	7p11.2	**1 (14); 2 (35)**	N. T.	1 (12); 2 (3	-1 (1)	**1 (24); 2 (1)**	-1 (2)	1 (5); 2 (5)	N. T.
SEC61G	7p11.2	**1 (16); 2 (32)**	N. T.	1 (12); 2 (3	-1 (1)	**1 (25)**	-1 (2)	1 (5); 2 (5)	N. T.
ITGB8	7p21.1	**1 (44); 2 (3)**	N. T.	1 (7); 2 (1)	N. T.	**1 (24); 2 (2)**	-1 (6); -2 (1)	1 (6); 2 (1)	N. T.
LFNG	7p22.3	**1 (43); 2 (2)**	N. T.	1 (13)	-1 (1)	**1 (23)**	-1 (9); -2 (1)	1 (6)	-1 (1)
PDGFA	7p22.3								
PMS2P11	7q11.23								
DENND2A	7q34	**1 (41); 2 (3)**	N. T.	1 (16)	N. T.	1 (27); 2 (12)	-1 (1)	1 (6); 2 (1)	N. T.
VAV3	1p13.3	1 (4)	-1 (2)	1 (3)	N. T.	1 (5)	**-1 (27); -2 (1)**	N. T.	-1 (2)
DMRTA2	1p33								
CDKN2C	1p33	1 (7)	-1 (3)	1 (2)	-1 (2); -2 (1)	1 (8)	**-1 (30)**	N. T.	-1 (2)
NFIA	1p31.3	1 (6)	N. T.	1 (2)	-2 (1)	1 (4)	**-1 (28)**	N. T.	-1 (2)
ZNF558	19p13.2	**1 (18)**	-1 (8)	1 (1)	-1 (5); -2 (1)	1 (8); 2 (2)	**-1 (23)**	1 (2)	-1 (10)
UHRF1	19p13.3	**1 (18)**	-1 (8)	1 (1)	-1 (5); -2 (1)	1 (8); 2 (2)	**-1 (23)**	1 (2)	-1 (10)
PRPF31	19q13.42	**1 (17)**	-1 (8)	1 (1)	-1 (5); -2 (1)	1 (5)	**-1 (44); -2 (5)**	1 (2)	-1 (10)
ACSS3	12q21.31	1 (3)	-1 (15)	1 (2)	-1 (7); -2 (1)	1 (5)	**-1 (41); -2 (3)**	N. T.	-1(7)
SOCS2	12q22	1 (3)	-1 (1)	1 (2)	-1 (3)	1 (1); 2 (1)	-1 (14)	N. T.	-1 (1)
HS3ST3B1	17p12	1 (2); 2 (1)	-1 (3)	N. T.	-1 (6)	1 (2)	-1 (15)	1 (1)	-1(4)
SOX9	17q24.3	1 (2)	-1 (3)	1 (1); 2 (1)	-1 (3)	1 (12)	-1 (7)	1 (1)	-1 (5)
NES	1q23.1	1 (6)	N. T.	1 (4); 2 (1)	N. T.	1 (10); 2 (2)	-1 (1)	N. T.	-1 (1)
SHOX2	3q25.32	1 (3)	N. T.	1 (1)	-1 (1)	1 (5)	-1 (6)	1 (1)	N. T.
RNF180	5q12.3	1 (3)	-1 (1)	1 (3)	N. T.	1 (2)	-1 (8); -2 (1)	1 (2)	N. T.
ELOVL2	6p24.2	N. T.	-1 (5)	N. T.	-1 (2)	1 (4)	-1 (7)	1 (1)	-1 (1)
POU3F2	6q16	N. T.	-1 (10)	N. T.	-1 (4)	1 (3)	-1 (5); -2 (2)	1 (1)	-1 (2)
TNFRSF19	13q12.12	N. T.	-1 (9)	1 (1)	-1 (6)	1 (2)	**-1 (27); -2 (3)**	N. T.	-1(4)
SEMA6D	15q21.1	N. T.	-1 (6)	1 (1)	-1 (1)	1 (3)	**-1 (18); -2 (1)**	N. T.	-1 (1)
SALL1	16q12.1	1 (1)	-1 (7)	1 (2)	-1 (6)	1 (5)	-1 (14)	N. T.	-1 (2)
SNX5	20p11	**1 (21)**	-1 (1)	1 (5)	N. T.	1 (12); 2 (1)	-1 (6); -2 (1)	N. T.	N. T.

Table S8. Recurrent and concomitant gain or loss of EM/PM claasifier genes in glioma genomes as analyzed in the REMBRANDT data set

NCOA3	20q12	**1 (20)**	-1 (1)	1 (5)	-1 (1)	1 (16)	-1 (3)	N. T.	-1 (1)
TRIOBP	22q13.1	1 (2)	**-1 (34)**	N. T.	-1 (13)	1 (9)	**-1 (32); -2 (1)**	1 (1)	-1(12)

PDGFRA-M	**Chr**	**EM+++ (n = 53)**		**EM++ (n = 28)**		**PM+++ (n =99)**		**EMlowPMlow (n = 25)**	
		Gain grade (n)	**Loss grade (n)**	**Gain grade (n**	**Loss grade (n)**	**Gain grade (n)**	**Loss grade (n)**	**Gain grade (n)**	**Loss grade (n)**
PDGFRA	4q12	1 (3)	N. T.	2 (1); 1 (2)	N. T.	1 (4); 2 (4)	-1 (13)	1 (1); 2 (1)	N. T.
LPHN3	4q13.1	1 (1)	N. T.	1 (1)	N. T.	1 (4); 2 (2)	-1(14)	1 (1)	N. T.
RP11-35N6.1 (LPPR1)	9q31.1								
ZNF462	9q31.2	N. T.	-1 (5)	1 (5)	-1 (1)	1 (11); 2 (1)	-1 (5)	N. T.	N. T.
C9orf140 (SAPCD2)	9q34.3	N. T.	-1 (9)	1 (5)	-1 (4)	1 (12); 2 (1)	-1(7); -2 (1)	N. T.	-1 (2); -2 (1)
LIX1L	1q21.1	1 (5)	N. T.	1 (4); 2 (1)	N. T.	1 (9); 2 (2)	-1 (3)	N. T.	-1 (1)
MEX3A	1q22								
FAM5C	1q31.1	1 (5); 2 (1)	-1 (1)	1 (4); 2 (1)	N. T.	1 (6); 2 (8)	-1 (1); -2 (1)	N. T.	N. T.
C1orf106	1q32.1	1 (7)	-1 (1)	1 (5); 2 (1)	N. T.	1 (10); 2 (1)	-1 (1); -2 (2)	N. T.	N. T.
NAV1	1q32.1	1 (7)	-1 (1)	1 (5); 2 (1)	N. T.	1 (10); 2 (1)	-1 (1); -2 (2)	N. T.	N. T.
OLIG1	21q22.11	1 (2)	-1 (6)	1 (2)	-1 (1)	1 (3)	-1(8); -2 (1)	N. T.	-1(3)
OLIG2	21q22.11	1 (2)	-1 (6)	1 (2)	-1 (1)	1 (3)	-1(8); -2 (1)	N. T.	-1(3)
PATZ1	22q12.2	1 (2)	**-1 (34)**	N. T.	-1 (12)	1 (9)	**-1(33)**	1 (1)	-1(12)
CSNK1E	22q13.1	1 (2)	**-1 (34)**	N. T.	-1 (13)	1 (9)	**-1(32); -2 (1)**	1 (1)	-1(12)
POLR2F	22q13.1	1 (2)	**-1 (34)**	N. T.	-1 (13)	1 (9)	**-1(32); -2 (1)**	1 (1)	-1(12)
PCGF2	17q12	1 (3)	-1 (3)	1 (1)	-1 (4)	1 (4)	-1(13)	1 (1)	-1(4)
C1QL1	17q21.31	1 (3)	-1 (3)	1 (1)	-1 (4)	1 (8)	-1 (9)	1 (1)	-1(4)
CACNG4	17q24.2	1 (2)	-1 (2)	1 (1); 2 (1)	-1 (3)	1 (11)	-1(3)	1 (1)	-1 (5)
TMCC1	3q22.1	1 (2)	-1 (3)	1 (1)	-1 (2)	1 (7)	-1 (5)	1 (1)	N. T.
NLGN1	3q26.31	1 (5)	N. T.	3 (1); 1 (1)	N. T.	1 (5); 2 (2)	-1 (2)	1 (1)	N. T.
SOX4	6p22.3	N. T.	-1 (5)	N. T.	-1 (3)	1 (4)	-1 (4)	1 (1)	-1 (1)
MARCKS	6q21	N. T.	**-1 (10); -2 (1)**	N. T.	-1 (4)	2 (1)	-1 (6); -2 (2)	1 (1)	-1 (2)
ETV1	7p21.2	**1 (43); 2 (2)**	N. T.	1 (14)	N. T.	**1 (22), 2(4)**	-1(7); -2 (1)	1 (6)	N. T.
NXPH1	7p21.3	**1 (43); 2 (2)**	N. T.	1 (14)	N. T.	**1 (22); 2(3)**	-1(8); -2 (1)	1 (6)	N. T.

Table S8. Recurrent and concomitant gain or loss of EM/PM claasifier genes in glioma genomes as analyzed in the REMBRANDT data set

CHD7	8q12.1	1 (2)	-1 (3)	2 (1)	N. T.	1 (15); 2 (3)	-1 (4); -2 (1)	N. T.	N. T.
MMP16	8q21.3	1 (2)	-1 (1)	2 (1)	N. T.	1 (16); 2 (7)	-1 (6)	1 (1)	N. T.
C10orf18 (FAM208B)	10p15.1	1 (3)	-1 (15)	2 (1)	-1 (7); -2 (1)	1 (5)	**-1(41); -2(3)**	N. T.	-1(7)
ZEB1	10p11.22								
ZNF22	10q11.21	1 (1)	**-1 (46)**	1 (1)	-1 (16)	1 (1)	**-1 (22)**	N. T.	-1(8)
EIF4EBP2	10q22.1	N. T.	**-1 (48)**	1 (1)	-1 (15)	1 (1)	**-1 (25)**	N. T.	-1(9); -2 (1)
TACC2	10q26.13	N. T.	**-1 (48); -2 (1)**	1 (1)	-1 (16)	N. T.	**-1 (29); -2 (5)**	N. T.	-1(8); -2 (2)
SOX6	11p15.2	1 (1)	-1 (3)	N. T.	-1 (1)	1 (8); 2 (1)	**-1 (16); -2 (6)**	N. T.	N. T.
LOC283174	11q25								
LOC100507187 (outside	11								
KLRC3	12p13.2	1 (3)	N. T.	1 (4)	-1 (1)	1 (6); 2 (4)	-1 (5); -2 (1)	N. T.	N. T.
NOVA1	14q12	N. T.	**-1 (11)**	1 (1)	-1 (7)	N. T.	-1 (13)	N. T.	-1 (1)
RFX7	15q21.3								
SOX8	16p13.3	1 (1)	-1 (6)	1 (1)	-1 (6)	1 (5)	-1 (11)	N. T.	-1(3)
TSHZ1	18q22.3	1 (2)	-1 (3)	1 (6)	-1 (3)	1 (4)	-1(13); -2 (1)	1 (1)	-1(3); -2 (1)
MYT1	20q13.33	**1 (20)**	-1 (1)	1 (6)	N. T.	1 (16)	-1 (3)	N. T.	N. T.
Other marker genes									
MET	7q31.2	1 (41); 1(3)	N. T.	1 (16)	N. T.	1 (25); 2 (16)	-1 (1)	1 (6); 2 (1)	N. T.
PTEN	10q23.31	N. T.	-1 (46); -2 (2)	N. T.	-1 (15)	N. T.	-1 (24); -2 (3)	N. T.	-1(8); -2 (2)
NF1	17q11.2	1 (3)	-1 (3)	N. T.	-1 (5)	1 (3)	-1 (14)	1 (1)	-1(4)
CDK6	7q21.2	1 (39); 2 (6)	N. T.	1(16)	N. T.	1 (26); 2 (12)	-1 (1)	1 (6); 2 (2)	N. T.
CDKN2A/2B	9p21.3	N. T.	-1 (14); -2 (24)	1 (2); 2 (1)	-1 (5); -2 (7)	N. T.	-1 (18); -2 (17)	N. T.	-1 (1); -2 (2)

The segmented copy number data of 100K SNP array data (*HindIII* part) from the REMBRANDT data set were analyzed using GISTIC 2.0 at a threshold of 0.2.
The extent of amplification (1, and 2) and deletion (-1 for heterozygous deletion, and -2 for homozygous loss) at a given gene locus in EM or PM gliomas is indicated.
N. T.: alteration not detected.
Concomitant gains or losses of genes are labelled with color. Bold labelled genes are enriched in developing astrocytes or OPCs as analyzed in Figure 5.
Focal amplification of EGFR and SEC61G were found in 66.0% and 62.3% of EM+++ gliomas, respectively. Twenty-four percent of EMlowPMlow gliomas are associated with focal amplification of EGFR/SEC61G.

Table S9. Co-occurrence of SCNAs of EM/PM classifiers in EM and PM gliomas of the TCGA mRNA-seq data set. Odds ratios and Fisher´s exact P values are presented.

EM Glioma (N=123)		Cytoband	EM GENE											
			EGFR	SEC61G	ITGB8	LFNG	PDGFA	DENND2A	CDKN2C	ZNF558	PRPF31	SNX5	NCOA3	TRIOBP
EM GENE	EGFR	7p11.2	N.A											
	SEC61G	7p11.2	123.30	N.A										
	ITGB8	7p21.1	41.40	4.90	N.A									
	LFNG	7p22.3	30.64	3.54	307.17	N.A								
	PDGFA	7p22.3	9.43	2.35	176.40	1852.00	N.A							
	DENND2A	7q34	41.40	14.29	25.47	41.78	24.37	N.A						
	CDKN2C	1p32.3	0.13	0.32	0.81	1.16	1.42	1.15	N.A					
	ZNF558	19p13.2	5.19	1.45	1.08	2.10	1.80	0.84	0.79	N.A				
	PRPF31	19q13.42	5.01	1.40	1.33	2.58	2.23	0.81	1.00	69.14	N.A			
	SNX5	20p11.23	3.93	1.08	1.31	1.92	1.66	1.00	0.58	4.63	5.82	N.A		
	NCOA3	20q13.12	3.93	1.08	1.31	1.49	1.29	1.00	0.46	4.63	5.82	799.50	N.A	
	TRIOBP	22q13.1	0.10	0.34	0.42	0.50	0.49	0.42	1.36	0.46	0.36	1.13	1.51	N.A

Odds ratio for the occurrence of the SCNAs in EM classifiers among 123 EM+++ gliomas.

<1 indicates tendency toward mutual exclusivity

1 indicates no association of gene alterations given the number present for each of the two genes

>1 indicates tendency toward co-occurrence

The corresponding P value less than 0.05

EM Glioma (N=123)		Cytoband	EM GENE											
			EGFR	SEC61G	ITGB8	LFNG	PDGFA	DENND2A	CDKN2C	ZNF558	PRPF31	SNX5	NCOA3	TRIOBP
EM GENE	EGFR	7p11.2	N.A											
	SEC61G	7p11.2	2.78E-04	N.A										
	ITGB8	7p21.1	0.004	0.038	N.A									
	LFNG	7p22.3	0.008	0.081	1.87E-18	N.A								
	PDGFA	7p22.3	0.089	0.367	6.49E-18	1.27E-24	N.A							
	DENND2A	7q34	0.004	6.00E-04	9.12E-10	3.83E-12	9.55E-10	N.A						
	CDKN2C	1p32.3	0.121	0.105	0.77	1	0.781	1	N.A					
	ZNF558	19p13.2	0.266	0.734	1	0.173	0.253	0.726	0.587	N.A				
	PRPF31	19q13.42	0.27	0.737	0.627	0.07	0.115	0.665	0.991	4.32E-18	N.A			
	SNX5	20p11.23	0.551	1	0.799	0.244	0.322	0.997	0.265	9.47E-05	9.21E-06	N.A		
	NCOA3	20q13.12	0.551	1	0.799	0.487	0.614	0.997	0.116	9.47E-05	9.21E-06	7.24E-25	N.A	
	TRIOBP	22q13.1	0.073	0.166	0.089	0.173	0.125	0.089	0.484	0.044	0.01	0.753	0.28	N.A

P values for the Odds ratio for the co-occurrence of the SCNAs in EM classifiers among 123 EM+++ gliomas.

P<0.05

Table S9. Co-occurrence of SCNAs of EM/PM classifiers in EM and PM gliomas of the TCGA mRNA-seq data set. Odds ratios and Fisher´s exact P values are presented.

EM Glioma	(N=123)	Cytoband	PM GENE				
			ETV1	NOVA1	PATZ1	CSNK1E	POLR2F
PM GENE	ETV1	7p21.2	N.A				
	NOVA1	14q12	0.65	N.A			
	PATZ1	22q12.2	0.43	3.20	N.A		
	CSNK1E	22q13.1	0.38	2.74	2022.00	N.A	
	POLR2F	22q13.1	0.45	2.58	1728.00	2889.00	N.A

Odds ratio for the occurrence of the SCNAs in PM classifiers among 123 EM+++ gliomas.

EM Glioma	(N=123)	Cytoband	PM GENE				
			ETV1	NOVA1	PATZ1	CSNK1E	POLR2F
PM GENE	ETV1	7p21.2	N.A				
	NOVA1	14q12	0.379	N.A			
	PATZ1	22q12.2	0.09	0.005	N.A		
	CSNK1E	22q13.1	0.055	0.017	5.34E-26	N.A	
	POLR2F	22q13.1	0.142	0.027	6.31E-26	7.76E-27	N.A

P values for the Odds ratio for the co-occurrence of the SCNAs in PM classifiers among 123 EM+++ gliomas.

PM Glioma	(N=133)	Cytoband	EM GENE		PM GENE				
			NFIA	PRPF31	ZEB1	EIF4EBP2	TACC2	MYT1	NOVA1
EM GENE	NFIA	1p31.3	N.A						
	PRPF31	19q13.42	30.96	N.A					
PM GENE	ZEB1	10p11.22			N.A				
	EIF4EBP2	10q22.1			22.95	N.A			
	TACC2	10q26.13			4.38	102.50	N.A		
	MYT1	20q13.33			9.35	0.00	0.00	N.A	
	NOVA1	14q12			0.79	0.00	0.00	0.00	N.A

Odds ratio for the co-occurrence of the SCNAs in EM/PM classifiers among 133 PM+++ gliomas.

PM Glioma	(N=133)	Cytoband	EM GENE		PM GENE				
			NFIA	PRPF31	ZEB1	EIF4EBP2	TACC2	MYT1	NOVA1
EM GENE	NFIA	1p31.3	N.A						
	PRPF31	19q13.42	1.01E-11	N.A					
PM GENE	ZEB1	10p11.22			N.A				
	EIF4EBP2	10q22.1			0.003	N.A			
	TACC2	10q26.13			0.012	1.93E-05	N.A		
	MYT1	20q13.33			0.002	0.672	0.443	N.A	
	NOVA1	14q12			1	0.76	0.375	0.571	N.A

P values for the Odds ratio of the occurrence of the SCNAs in EM/PM classifiers among 133 PM+++ gliomas.

Table S10. Regression analysis between the expression of EM/PM classifiers and EM/PM signature.

		TCGA mRNA-seq data set				REMBRANDT data set				GSE16011 data set			
		EM+++(123) regression coefficient		PM+++(133) regression coefficient		EM gliomas (149) regression coefficient		PM gliomas(230) regression coefficient		EM gliomas(86) regression coefficient		PM gliomas(52) regression coefficient	
EGFR-M	**Chr**	**β**	**p value**	**β**	**p value**	**β**	**p value**	**β**	**p value**	**β**	**p value**	**β**	**p value**
EGFR	7p11.2	0.052	<0.0001	0.036	0.002	0.149	<0.0001	0.067	<0.0001	0.505	<0.0001	No Effect	
SEC61G	7p11.2	0.036	<0.0001	No Effect		0.047	<0.0001	0.028	<0.0001	0.047	<0.0001	0.029	0.019
ITGB8	7p21.1	0.115	0.004	0.117	<0.0001	0.355	<0.0001	0.19	<0.0001	No Effect		1.155	<0.0001
LFNG	7p22.3	0.313	<0.0001	0.189	<0.0001	0.393	<0.0001	0.263	<0.0001	No Effect		0.451	<0.0001
PDGFA	7p22.3	0.484	<0.0001	No Effect		0.545	<0.0001	No Effect		0.701	0.004	No Effect	
DENND2A	7q34	0.381	<0.0001	0.444	<0.0001	0.532	<0.0001	0.237	<0.0001	0.576	0.001	0.44	<0.0001
CDKN2C	1p33	0.355	0.006	0.15	0.04	0.484	<0.0001	0.058	<0.0001	0.219	0.019	No Effect	
NFIA	1p31.3	No Effect		0.17	<0.0001	0.327	<0.0001	0.131	<0.0001	0.185	0.004	0.101	<0.0001
ZNF558	19p13.2	2.727	<0.0001	No Effect		0.455	0.004	No Effect		0.643	<0.0001	0.09	0.042
PRPF31	19q13.42	1.227	<0.0001	0.429	<0.0001	1.145	<0.0001	0.312	<0.0001	0.98	0.001	0.423	<0.0001
SNX5	20p11	No Effect		0.26	<0.0001	0.644	<0.0001	0.348	<0.0001	0.467	<0.0001	0.114	0.001
NCOA3	20q12	1.488	0.013	No Effect		1.758	<0.0001	0.517	<0.0001	1.597	0.014	No Effect	
TRIOBP	22q13.1	1.541	0.002	0.452	0.014	1.664	<0.0001	0.268	0.018	1.901	0.002	No Effect	
PDGFRA-M													
PDGFRA	4q12	0.036	<0.0001	0.046	<0.0001	0.071	<0.0001	0.106	<0.0001	0.057	<0.0001	0.042	<0.0001
PATZ1	22q12.2	0.906	<0.0001	1.652	<0.0001	1.964	<0.0001	2.021	<0.0001	4.828	<0.0001	4.487	<0.0001
CSNK1E	22q13.1	0.408	<0.0001	0.399	<0.0001	0.794	<0.0001	0.657	<0.0001	1.427	<0.0001	1.077	0.001
POLR2F	22q13.1	No Effect		0.619	<0.0001	0.685	<0.0001	0.41	<0.0001	0.303	<0.0001	0.221	<0.0001
ETV1	7p21.2	0.203	<0.0001	0.121	<0.0001	0.209	<0.0001	0.264	<0.0001	0.302	<0.0001	0.315	<0.0001
ZEB1	10p11.22	0.365	<0.0001	0.304	<0.0001	0.266	<0.0001	0.316	<0.0001	0.271	<0.0001	0.274	<0.0001
EIF4EBP2	10q22.1	No Effect		0.461	<0.0001	0.557	0.02	1.2	<0.0001	No Effect		No Effect	
TACC2	10q26.13	1.064	<0.0001	1.586	<0.0001	0.685	<0.0001	0.681	<0.0001	0.8	0.001	0.989	<0.0001
NOVA1	14q12	0.481	<0.0001	1.009	<0.0001	0.298	<0.0001	0.295	<0.0001	0.437	<0.0001	0.366	0.002
MYT1	20q13.33	1.268	<0.0001	0.851	<0.0001	1.172	<0.0001	0.628	<0.0001	1.377	<0.0001	0.752	<0.0001

Linear regression analysis between the expression of EM classifiers and EM signature, PM classifiers and PM signature was first performed in the TCGA mRNA-seq cohort, and subsequently replicated in the REMBRANDT and GSE16011 data sets. Regression coefficient (β values) and their p values are indicated.
Genes with high regression coefficients between its own expression level and the entire EM or PM signature are shown in bold script.

Table S11. Regional chromosomal alterations in PM+++ and EM+++ gliomas from the TCGA mRNA-seq data base.

PM+++ (n=133)

Amplification	q value	residual q valu	wide peak boundaries	frequency	genes in wide peak
7p11.2	7.22E-72	1.81E-69	chr7:54380374-54394400	32.30%	[HPVC1]
1q21.3	9.01E-54	1.61E-39	chr1:152571938-152579824	37.60%	LCE3C
5q11.2	3.49E-42	3.39E-38	chr5:57329848-57332553	30.10%	[PLK2]
20p13	9.29E-38	9.29E-38	chr20:1570738-1571902	40.60%	SIRPB1
17q21.2	9.03E-41	3.27E-36	chr17:39423211-39427849	43.60%	[KRTAP9-9]
4q34.1	5.49E-42	1.74E-35	chr4:172374950-172377457	45.50%	[GALNTL6]
8q23.3	1.61E-19	3.47E-16	chr8:112294467-112301251	30.90%	[CSMD3]
4p16.1	1.06E-19	3.51E-16	chr4:10228784-10229275	46.60%	[WDR1]
1p31.1	3.40E-21	1.32E-14	chr1:72773430-72789610	48.00%	[NEGR1]
8p23.3	2.12E-24	2.19E-14	chr8:594974-606116	30.90%	[ERICH1]
3q29	1.54E-20	5.06E-14	chr3:192877963-192880586	33.10%	[HRASLS]
9q13	7.37E-12	2.56E-11	chr9:68183979-68188403	18.00%	[ANKRD20A3]
6p21.32	5.98E-17	3.45E-11	chr6:32518770-32544308	36.80%	HLA-DRB6
4p16.1	1.19E-14	1.91E-09	chr4:9451924-9478848	21.00%	DEFB131
10p12.2	2.89E-11	2.57E-09	chr10:24377681-24377920	27.10%	KIAA1217
2q22.3	8.23E-11	4.42E-09	chr2:146865328-146866706	33.80%	[PABPC1P2]
8p11.22	2.09E-12	1.52E-08	chr8:39247112-39352686	39.10%	ADAM3A ADAM5P
3p14.1	6.85E-09	1.77E-08	chr3:68746260-68748785	27.10%	[FAM19A4]
16q23.1	4.82E-09	8.67E-08	chr16:78376378-78384737	26.30%	WWOX
7q34	3.16E-13	9.98E-08	chr7:142474940-142479806	41.30%	TRY6

Deletion	q value	residual q valu	wide peak boundaries		genes in wide peak
20p13	1.68E-66	1.68E-66	chr20:1536809-1600924	34.60%	SIRPB1
1q21.3	7.69E-74	1.69E-49	chr1:152552820-152586539	39.10%	LCE3C
11p15.4	7.34E-57	5.14E-49	chr11:5776624-5809138	48.80%	OR52N5
1p13.3	4.46E-62	4.86E-46	chr1:110226488-110256623	53.40%	GSTM1

22q11.23	1.89E-48	7.47E-44	chr22:24324874-24382761	13.50%	GSTTP1	LOC391322	
15q24.3	3.44E-54	3.72E-43	chr15:76632488-77227699	32.30%	SCAPER		
14q24.3	2.11E-53	6.96E-43	chr14:73992773-74017520	31.50%	ACOT1		
7q34	1.76E-54	2.82E-41	chr7:142455689-142483350	21.80%	PRSS1	PRSS2	TRY6
11q11	1.61E-35	6.11E-34	chr11:55363341-55541283	33.10%	OR4P4, OR4C11, OR4S2, OR4C6		
17p11.2	1.64E-34	8.27E-33	chr17:18319863-18426460	24.10%	USP32P2	LOC339240	LGALS9C
5q35.3	1.00E-38	8.64E-32	chr5:180374524-180467481	21.80%	BTNL3		
8p11.22	1.86E-45	3.42E-29	chr8:39256740-39422050	21.80%	ADAM3A		
8p23.1	4.20E-33	3.94E-27	chr8:12162783-12260972	24.60%	DEFB109P1, DEFB130, LOC100133267		
16q12.2	1.22E-34	3.18E-26	chr16:55782285-55842353	26.30%	CES1P1		
12p13.2	3.17E-40	1.15E-25	chr12:11236837-11260274	3.80%	TAS2R43		
19q13.41	1.57E-40	1.15E-25	chr19:52131814-52198653	62.40%	hsa-mir-125a, MIRLET7E, MIR125A, SIGLEC14		
10p12.2	1.01E-31	2.68E-25	chr10:23730192-24837906	1.50%	hsa-mir-603, KIAA1217, MIR603, PRINS		
19q13.12	2.42E-30	6.83E-23	chr19:35840043-35896967	50.40%	FFAR1, FFAR3		
1p36.11	1.78E-46	1.25E-22	chr1:25571270-25673153	42.90%	RHD		
6p12.1	3.31E-31	1.62E-21	chr6:53787268-54179039	1.50%	MLIP		

deletion peaks related to 1p19q loss

1p36.21	chr1:12883C	8.58E-12	0.028946	**46.60%**
1p36.13	chr1:170059	1.81E-14	2.83E-06	**49.60%**
1p36.12	chr1:223119	7.44E-05	0.069345	42.90%
1p36.11	chr1:255712	1.78E-46	1.25E-22	42.90%
19q13.42	chr19:55323	9.00E-15	0.013902	59.40%
19q13.41	chr19:52131	1.57E-40	1.15E-25	62.40%
19q13.31	chr19:43703	1.35E-25	6.14E-14	60.90%
19q13.12	chr19:3584C	2.42E-30	6.83E-23	50.40%
9p21.3	**chr9:21959C**	**2.10E-08**	**1.83E-07**	**30.10%**

EM+++ (n=123)

amplification	q value	residual q valu	wide peak boundaries		genes in wide peak		
7p11.2	1.20E-108	1.24E-108	chr7:55048002-55075819	96.70%	[**EGFR**]		
5q11.2	0.0044133	0.0044133	chr5:57321587-57332553	43.10%	[PLK2]		
12q15	6.01E-06	0.011125	chr12:69109665-69120901	10.60%	NUP107		
20p13	0.011125	0.011125	chr20:1570787-1571324	43.90%	SIRPB1		
7q21.2	0.017145	0.017145	chr7:92496797-92499294	88.60%	[**CDK6**]		
12q14.1	0.0002892	0.017145	chr12:59171211-59195089	8.10%	[LRIG3]		
17q21.2	0.017145	0.017145	chr17:39413411-39428141	39.80%	[KRTAP9-9]		
1p31.1	0.017145	0.020391	chr1:72788726-72804737	52.00%	[NEGR1]		
1q21.3	0.017145	0.05155	chr1:152571170-152579467	40.60%	LCE3C		
1q32.1	0.017145	0.05155	chr1:204436182-204488590	27.60%	**MDM4**	**PIK3C2B**	
2p16.3	0.05155	0.05155	chr2:52758791-52759031	46.30%	[ASB3]		
3q29	0.05155	0.05155	chr3:192877903-192882902	33.30%	[HRASLS]		
3p14.1	0.05155	0.05155	chr3:68746709-68748757	26.00%	[FAM19A4]		
4p16.1	0.017145	0.05155	chr4:10228881-10229291	39.00%	[WDR1]		
4q34.1	0.051543	0.05155	chr4:172374950-172377287	27.60%	[GALNTL6]		
5p11	0.051543	0.05155	chr5:46271009-46278211	25.20%	[HCN1]		
5p15.2	0.05155	0.05155	chr5:12815866-12833856	39.80%	[TAG]		
6p21.32	0.051543	0.05155	chr6:32501300-32530143	35.00%	HLA-DRB6		
8p23.3	0.05155	0.05155	chr8:595444-606116	23.60%	[ERICH1]		
8p11.22	0.05155	0.05155	chr8:39235604-39256500	39.00%	ADAM5P		

deletion	q value	residual q valu	wide peak boundaries		genes in wid		
9p21.3	9.15E-175	9.15E-175	chr9:21959052-21977193	79.70%	CDKN2A	C9orf53	
10q11.22	2.09E-69	9.84E-63	chr10:47087831-47142190	86.20%	LOC728643		
1q21.3	1.72E-79	1.77E-61	chr1:152552820-152586539	47.10%	LCE3C		
22q11.23	5.79E-54	1.94E-49	chr22:24371071-24385545	47.10%	GSTT1	LOC391322	
20p13	1.97E-45	1.97E-45	chr20:1536809-1600924	43.90%	SIRPB1		
1p13.3	3.19E-43	9.06E-34	chr1:110226488-110256506	43.10%	GSTM1		
7q34	1.58E-31	8.71E-27	chr7:142455689-142483411	21.10%	PRSS1	PRSS2	TRY6
11p15.4	1.55E-28	6.69E-24	chr11:5799831-5841855	19.50%	OR52N1		

11q11	4.41E-24	3.25E-23	chr11:55363341-55419817	29.30%	OR4P4	OR4C11	OR4S2
17p11.2	2.56E-23	1.35E-22	chr17:18426462-18540598	9%	FAM106A	CCDC144B	
3q22.1	1.58E-26	9.08E-22	chr3:129695715-129819637	4.90%	ALG1L2		
5q35.3	5.07E-23	9.08E-22	chr5:180374524-180467481	22.80%	BTNL3		
10p15.1	3.38E-26	2.03E-21	chr10:5557042-5726820	95.10%	CALML3	ASB13	
15q24.3	1.02E-33	2.66E-20	chr15:76631798-77227699	6.50%	SCAPER		
8p23.1	3.85E-24	2.30E-19	chr8:12162783-12260972	25.20%	DEFB109P1, DEFB130, LOC100133267		
14q24.3	1.02E-28	4.11E-17	chr14:73992773-74017520	37.40%	ACOT1		
4q13.2	4.80E-24	1.25E-16	chr4:70076939-70346056	40.60%	UGT2B28		
8q13.3	1.35E-16	3.88E-16	chr8:71645688-72759772	6.50%	EYA1	MSC	
16q12.2	7.48E-24	1.18E-15	chr16:55782285-55842353	33%	CES1P1		
15q11.2	6.40E-21	3.62E-15	chr15:22042519-22375469	30.90%	hsa-mir-3118		

other deletion peaks according to current notion

10q23.31	**6.03E-11**	**6.77E-07**	**chr10:89617158-90034038(pro**	**97.60%**
1p36.33	2.08E-05	0.17774	chr1:1627919-1687649(probes	14.60%
1p36.33	4.76E-12	0.004027	chr1:1-228705(probes 1:18)	26.00%
1p36.23	7.76E-16	1.85E-06	chr1:7908902-8386729(probes	29.20%
19q13.42	6.02E-07	3.18E-05	chr19:55323094-55347176(pro	15.40%
19q13.41	6.57E-16	1.81E-12	chr19:52131814-52198653(pro	24.30%
19q13.31	2.56E-05	0.096586	chr19:43703505-43766658(pro	16.30%
19q13.12	7.39E-10	1.97E-06	chr19:35840043-35896967(pro	18.70%
13q14.3	8.71E-09	1.12E-06	chr13:53050177-53194688(pro	33.30%

The top 20 amplified peaks according to the residual q value were presented.

The SNP6.0 data from the TCGA mRNA-seq data set were analyzed by GISTIC2.0 at an amplitude threshold of ± 0.2 under the supervison of EM/PM classification. Significant regional amplifications/deletions are shown. Human genome build hg19 was used as the reference.

Genes localized in the chromosomal regions close to the amplified or deleted wide peaks are indicated in bracket.

Table S12. Regional chromosomal alteration in EM/PM gliomas from the REMBRANDT data set.

PM+++ (n=99)

Amplification	wide peak boundaries	q value	residual q \	Frequency	# of genes in peak (known or interesting candidates)
2p24.3	chr2:15759534-16359887	0.018542	0.018542	11.10%	MYCN
8q24.21	chr8:128477299-131099761	0.021294	0.021294	32.30%	4 (MYC)
12q14.1	chr12:55667943-56745332	0.021294	0.021294	7.00%	39 (CDK4)
12q15	chr12:68177830-68277941	0.023281	0.034987	5.10%	2
12q15	chr12:67295984-67955712	0.023281	0.036352	6.10%	6 (MDM2)
12q14.1	chr12:60524779-60993852	0.084964	0.10407	4.00%	2
12q23.2	chr12:101924363-102483337	0.2115	0.2115	3.00%	1
1q31.1	chr1:182886426-183111389	0.24607	0.24607	14.10%	5
4p15.1	chr4:34581923-35275342	0.24607	0.24607	7.10%	1
12p13.33	chr12:546378-1636193	0.24607	0.24607	14.10%	7
Deletion					
9p21.3	chr9:21456646-22474015	2.18E-24	8.64E-22	35.30%	5 (CDKN2A/2B)
1p36.32	chr1:3788582-5940837	1.01E-06	1.01E-06	44.00%	3
9p24.3	chr9:1-316129	2.46E-09	5.20E-06	27.30%	3
13q14.2	chr13:38442698-59193226	2.71E-05	2.71E-05	33.30%	86 (RB1)
19q13.41	chr19:55514221-58685254	0.0002	0.000249	50.50%	87
5q35.2	chr5:174888401-180857866	0.0007	0.000674	21.20%	87 (FGFR4)
11p15.4	chr11:1-15964810	0.0007	0.000756	25.20%	249
14q24.3	chr14:63956180-106368585	0.0010	0.001021	24.20%	281 (JAG2, PSEN1, NUMB)
4q35.1	chr4:149670175-191411218	0.0016	0.001571	27.30%	128
10q26.3	chr10:120480890-135413628	0.0016	0.001583	35.40%	94
8p23.3	chr8:1-794884	0.0142	0.014335	12.10%	6
6q23.3	chr6:114462386-170975699	0.0273	0.026786	13.10%	244
5p15.33	chr5:1-2028460	0.0297	0.029658	13.10%	24
12q14.3	chr12:65355271-66871873	0.0594	0.060156	17.20%	3 (IFNG)
15q21.2	chr15:33600671-57188190	0.0959	0.094744	18.20%	179 (DLL4)
6p25.3	chr6:1-13804543	0.2404	0.25274	9.10%	70 (ELOVL2)

EM+++ (N=53)

Amplification	wide peak boundaries	q value	residual q \	Frequency	# of genes in peak (known or interesting candidates)
7p11.2	chr7:54667627-54753989	4.44E-105	4.63E-105	92.40%	SEC61G (EGFR is also amplified at 7p11.2)
7q21.2	chr7:91806694-92389778	0.0016	0.0018	84.90%	4 (CDK6)
12q14.1	chr12:56263308-56745610	0.0180	0.0180	13.20%	18 (CDK4)
12q15	chr12:69233208-69562460	0.0188	0.0188	11.30%	2 (PTPRB, PTPRR)
1q32.1	chr1:200569867-202083359	0.1967	0.1967	18.90%	21 (MDM4)

Deletion	wide peak boundaries	q value	residual q value		
9p21.3	chr9:21456646-22071849	6.02E-62	6.82E-53	71.70%	4 (CDKN2A/2B)
10q23.31	chr10:89677469-90434738	4.82E-06	4.82E-06	90.60%	3
9p21.3	chr9:22418154-25764341	4.91E-30	0.0001	71.60%	3
1p36.23	chr1:7704399-8402637	0.0003	0.0003	30.20%	7
6q27	chr6:167420583-170975699	0.0022	0.0278	26.40%	25 (DLL1)
6q22.1	chr6:101420223-161877726	0.0078	0.0503	24.50%	262
14q11.2	chr14:1-44721430	0.1024	0.1040	20.80%	189
6p25.3	chr6:1-25419439	0.1480	0.1480	15.10%	118
1p22.2	chr1:78826516-99046596	0.2172	0.2172	11.30%	83

EM++ (N=28)

Deletion	wide peak boundaries	q value	residual q \	Frequency	# of genes in peak (known or interesting candidates)
9p21.3	chr9:21456646-23697614	6.30E-14	6.30E-14	42.90%	5 (CDKN2A)
1p36.33	chr1:1-9898973	0.01496	0.0150	28.60%	114
19q13.31	chr19:49659555-55512606	0.01496	0.0150	21.40%	186
13q14.2	chr13:47287584-50212763	0.02515	0.0251	28.60%	21 (RB1)
16p12.1	chr16:21565891-27037655	0.06117	0.0625	35.70%	32
19p13.3	chr19:1-6912181	0.09660	0.0966	21.40%	198
20p12.3	chr20:5043138-5769184	0.11623	0.1162	7.10%	4
1q43	chr1:232938059-233207743	0.16015	0.1556	7.10%	2
7p11.2	chr7:54575924-55268216	0.15560	0.1556	7.10%	2 (EGFR, SEC61G)

EMlowPMlow (N=25)

Amplification	wide peak boundaries	q value	residual q v	Frequency	# of genes in peak (known or interesting candidates)
7p11.2	chr7:54492535-54754554	1.63E-05	1.63E-05	44.00%	SEC61G (EGFR is also amplified at 7P11.2)
12q14.1	chr12:56263308-57003704	0.011623	0.0116	12.00%	18 (CDK4)

Deletion	wide peak boundaries	q value	residual q v	Frequency	# of genes in peak (known or interesting candidates)
1p36.33	chr1:1-35113129	0.0011	0.0011	28.0%	437
9p21.3	chr9:19744477-23697614	0.0062	0.0062	12.0%	24 (CDKN2A)
10p11.21	chr10:35054441-35711156	0.0173	0.0173	32.0%	2
18q23	chr18:71260683-76117153	0.0622	0.0622	20.0%	14
6q13	chr6:64474951-170975699	0.1040	0.1040	12.0%	402
9q34.11	chr9:128416944-138429268	0.1040	0.1040	12.0%	173
7q11.21	chr7:56502097-68530168	0.1568	0.1568	20.0%	19
10q24.2	chr10:50710370-135413628	0.1590	0.1568	44.0%	490
16q22.3	chr16:73491781-73901976	0.1568	0.1568	8.0%	6
8p23.2	chr8:1981170-6259563	0.1717	0.1718	8.0%	1
17q24.1	chr17:61450204-61828085	0.1718	0.1718	24.0%	1
7q21.12	chr7:86647718-87480119	0.2580	0.2580	4.0%	6
16p13.3	chr16:4094853-6121015	0.2338	0.2580	12.0%	24

The 50K *HindIII* SNP array data from the REMBRANDT data set were analyzed by GISTIC2.0 at an amplitude threshold of ± 0.2 under the supervison of EM/PM classification. Significant regional amplifications/deletions are shown. Human genome build hg17 was used as the reference.

Table S13
GSEA analysis for differential enrichment of developing astrocyte or oligodendrocyte signatures in EM/PM gliomas.

Data set	Phenotype	Classifier	Size	NES	NOM p-val	FDR q-val	RANK at max
Tiantan							
	EM+++ vs PM+++	Developing Astro	303	-1.761	0.016	0.014	2635
	EM+++ vs PM+++	Developing Oligo	183	1.916	0.008	0.006	3018
GSE16011							
	EM+++ vs PM+++	Developing Astro	287	-1.526	0.008	0.012	2362
	EM+++ vs PM+++	Developing Oligo	161	1.460	0.039	0.031	1398
REMBRANDT							
	PM+++ vs EM+++	Developing Astro	306	-1.403	0.0400	0.0730	3284
	PM+++ vs EM++	Developing Astro	306	-1.329	0.0840	0.1140	2915
	PM+++ vs EMlowPMlow	Developing Astro	306	-0.880	0.6404	0.6330	3484
	PM+++ vs EM+++	Developing Oligo	184	1.496	0.0120	0.0212	1873
	PM+++ vs EM++	Developing Oligo	184	1.510	0.0249	0.0162	1909
	PM+++ vs EMlowPMlow	Developing Oligo	184	-1.338	0.1548	0.2643	2667