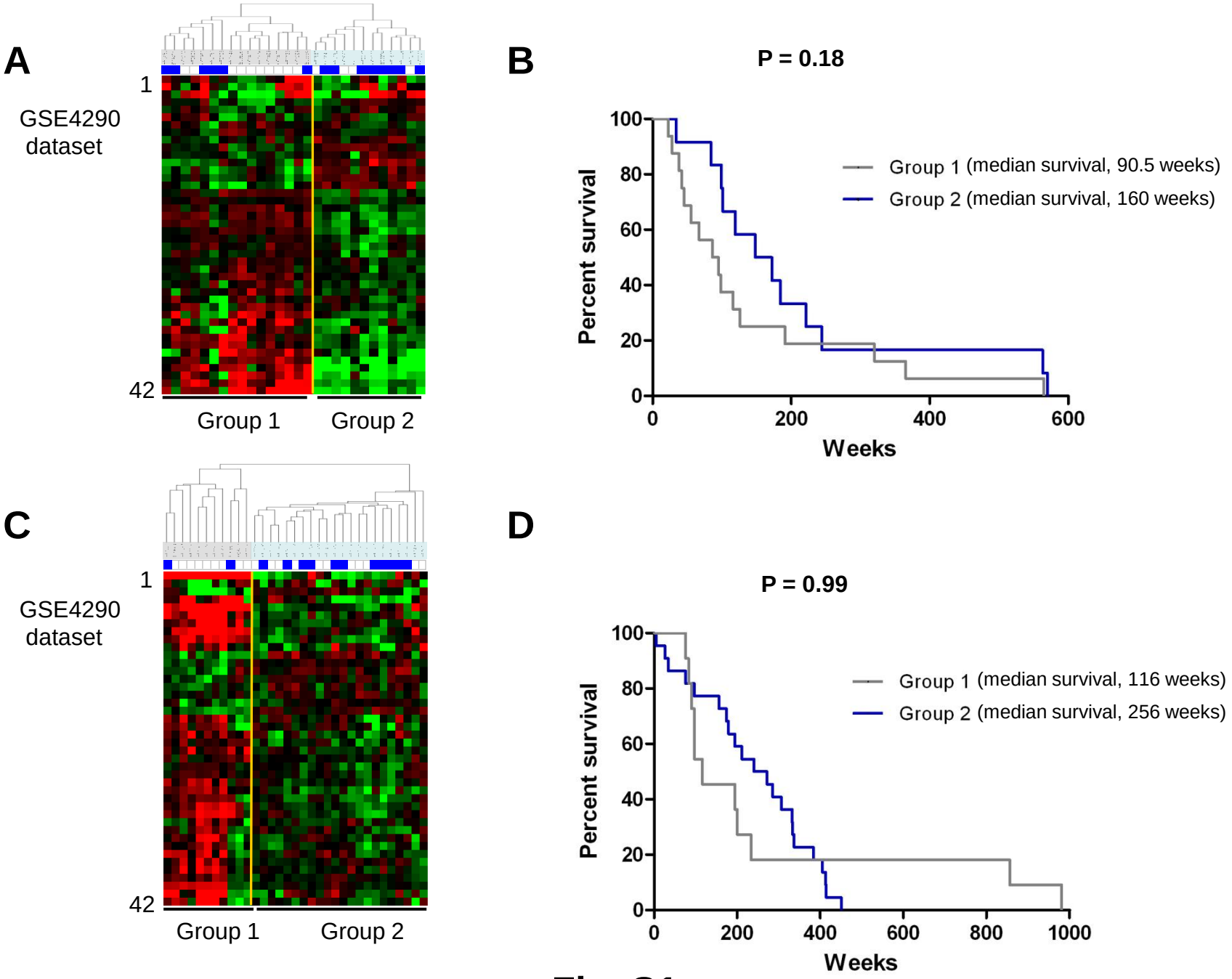
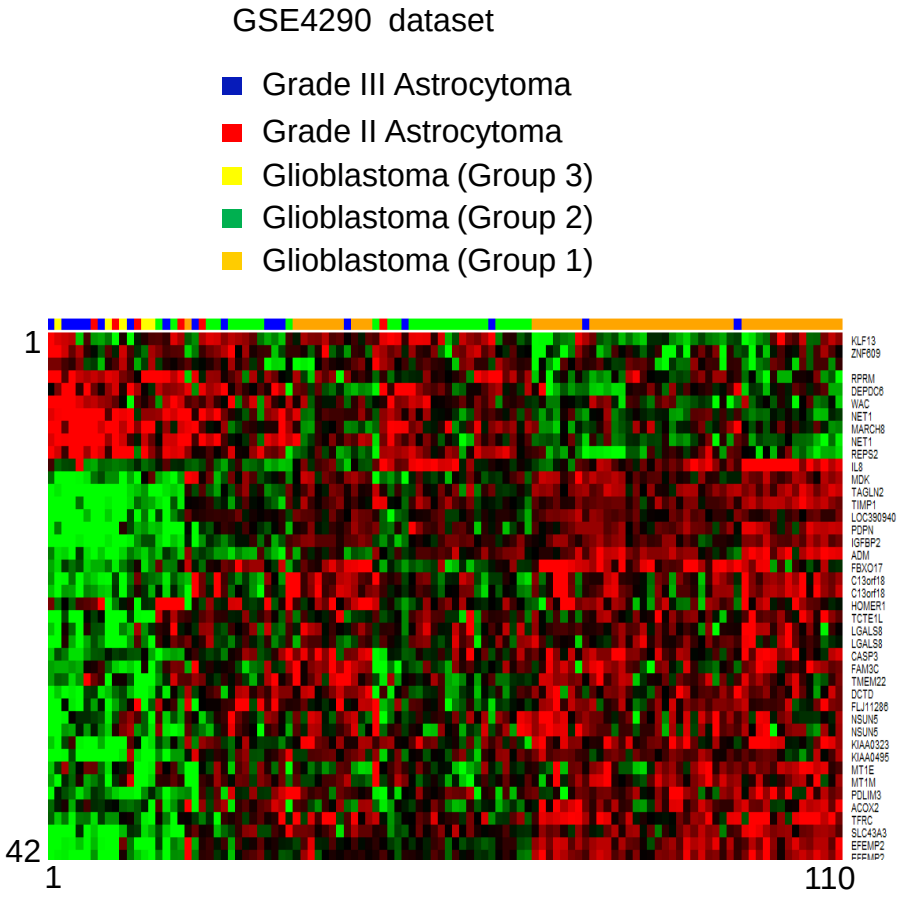

A
GSE4290
dataset
1
42
Group 1
Group 2
B
P = 0.18
Percent survival
100
80
60
40
20
0
0
200
400
600
Weeks
Group 1 (median survival, 90.5 weeks)
Group 2 (median survival, 160 weeks)
C
GSE4290
dataset
1
42
Group 1
Group 2
D
P = 0.99
Percent survival
100
80
60
40
20
0
0
200
400
600
800
1000
Weeks
Group 1 (median survival, 116 weeks)
Group 2 (median survival, 256 weeks)

**Fig. S1**


GSE4290 dataset
Grade III Astrocytoma
Grade II Astrocytoma
Glioblastoma (Group 3)
Glioblastoma (Group 2)
Glioblastoma (Group 1)
1
42
1
110
KLF13
ZNF609
RPRM
DEPDC6
WAC
NET1
MARCH8
NET1
REPS2
IL8
MDK
TAGLN2
TIMP1
LOC390940
PDPN
IGFBP2
ADM
FBXO17
C13orf18
C13orf18
HOMER1
TCTE1L
LGALS8
LGALS8
CASP3
FAM3C
TMEM22
DCTD
FLJ11286
NSUN5
NSUN5
KIAA0323
KIAA0495
MT1E
MT1M
PDLIM3
ACOX2
TFRC
SLC43A3
EFEMP2

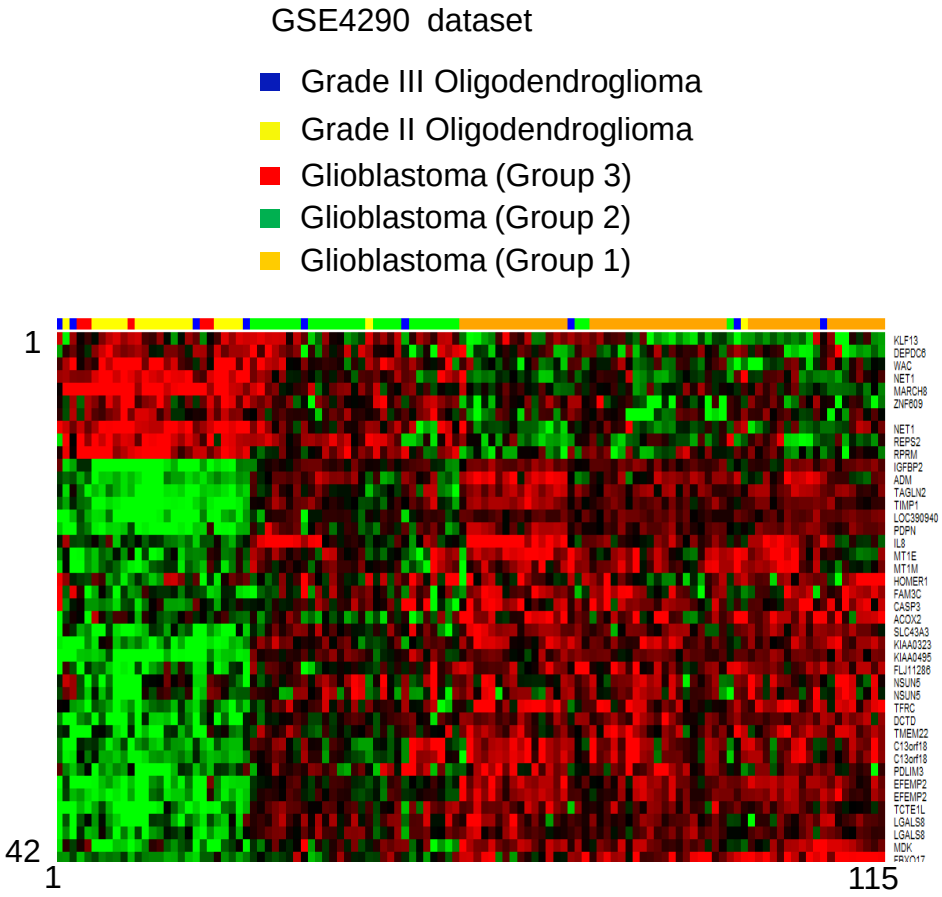
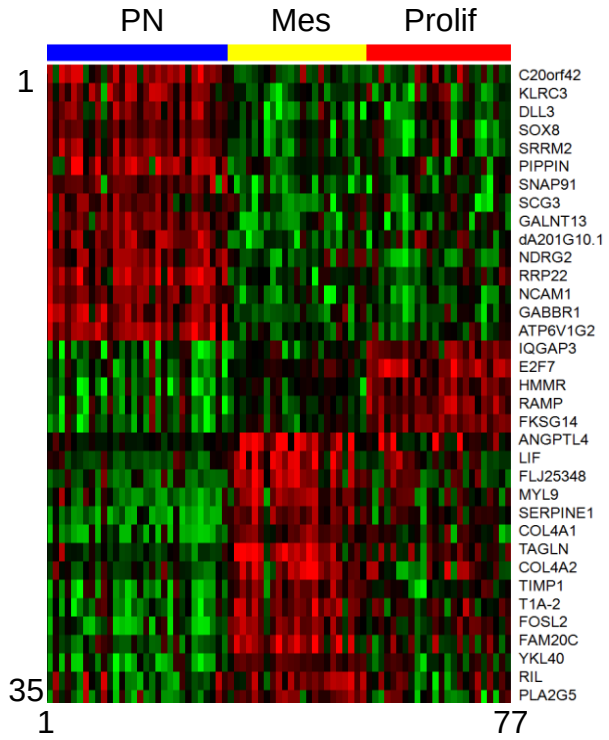

GSE4290 dataset
Grade III Oligodendroglioma
Grade II Oligodendroglioma
Glioblastoma (Group 3)
Glioblastoma (Group 2)
Glioblastoma (Group 1)
1
42
1
115
KLF13
DEPDC6
WAC
NET1
MARCH8
ZNF609
NET1
REPS2
RPRM
IGFBP2
ADM
TAGLN2
TIMP1
LOC390940
PDPN
IL8
MT1E
MT1M
HOMER1
FAM3C
CASP3
ACOX2
SLC43A3
KIAA0323
KIAA0495
FLJ11286
NSUN5
NSUN5
TFRC
DCTD
TMEM22
C13orf18
C13orf18
PDLIM3
EFEMP2
EFEMP2
TCTE1L
LGALS8
LGALS8
MDK

**Fig. S2**

**A** Phillips et al. 2006: using 35 signature genes




**B** Phillips et al. 2006: using 41 probesets (genes)

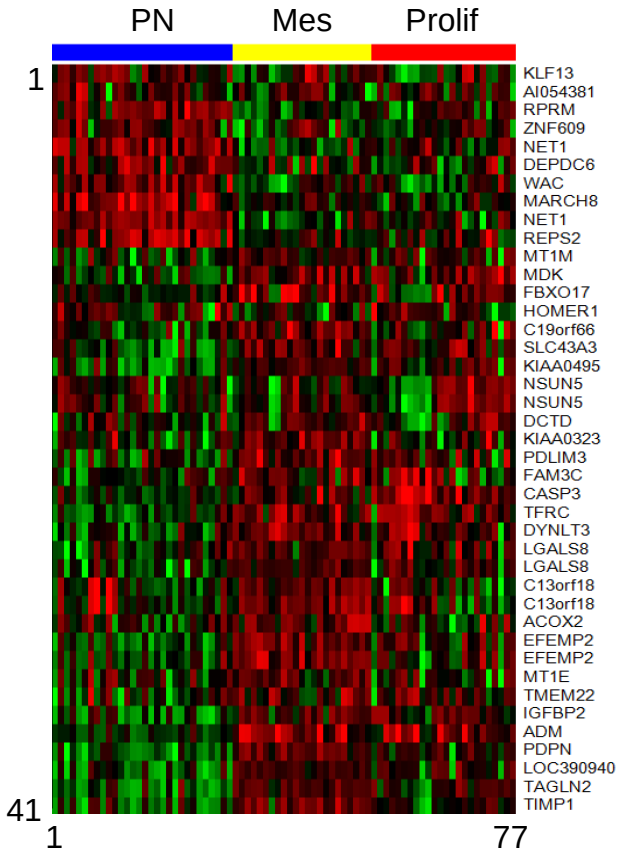
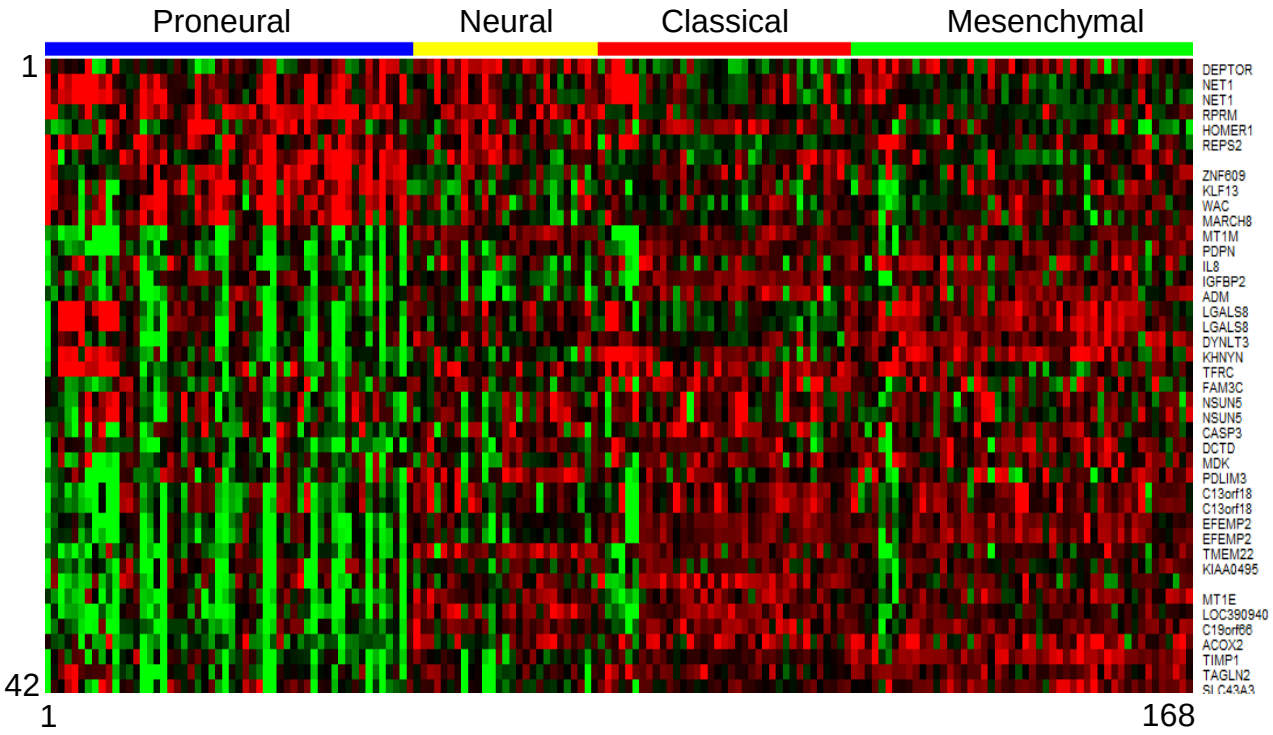



**Fig. S3**

**A**

Verhaak et al. 2010: using 42 probesets in TCGA Core samples




**B**

| Row Labels | Proneural | Neural | Classical | Mesenchymal | Grand Total |
|---|---|---|---|---|---|
| Group 1 | 13 | 13 | 37 | 41 | 104 |
| Group 2 | 17 | 7 | 11 | 13 | 48 |
| Group 3 | 14 | 2 | 0 | 0 | 16 |
| Grand Total | 44 | 22 | 48 | 54 | 168 |

**Fig. S4**

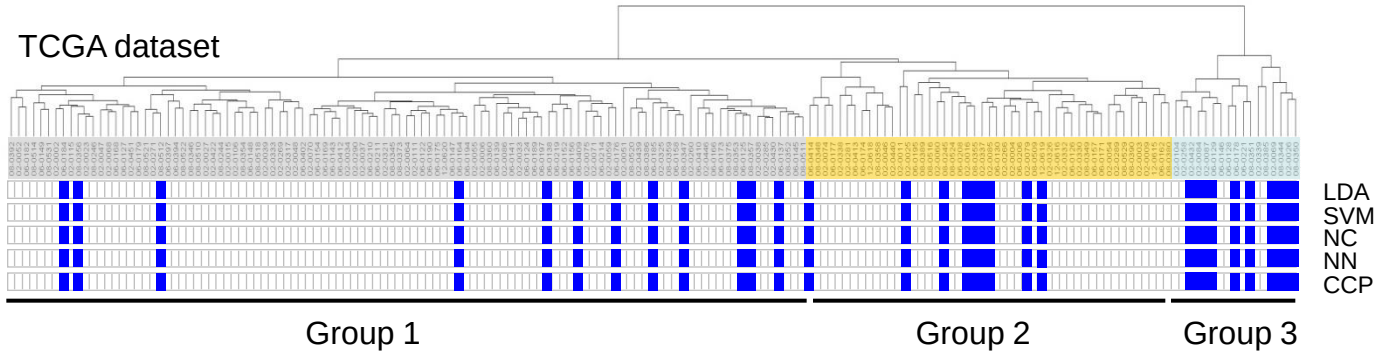



**Fig. S5**

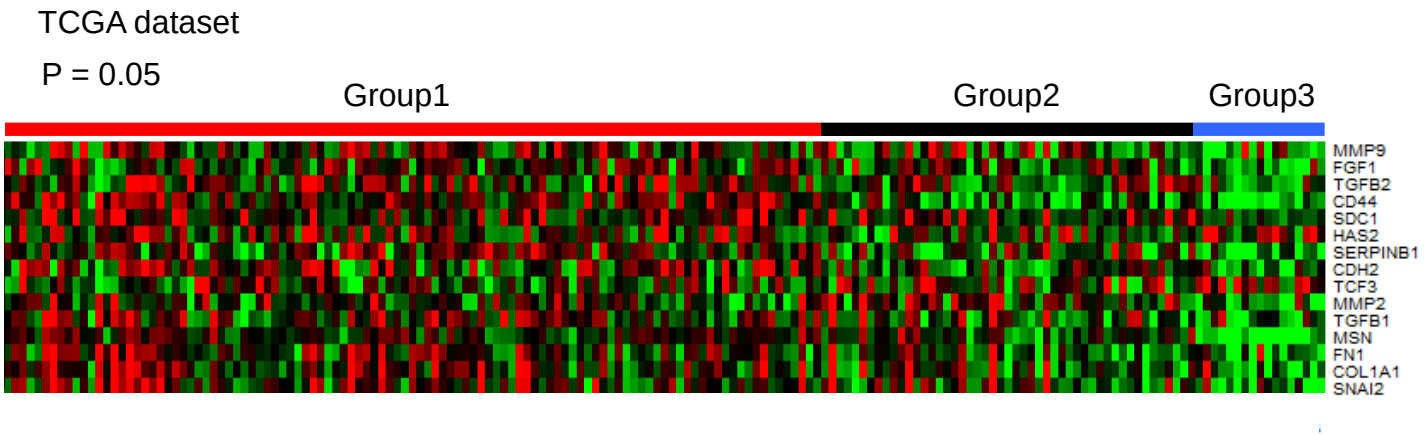

TCGA dataset
P = 0.05
Group1
Group2
Group3
MMP9
FGF1
TGFB2
CD44
SDC1
HAS2
SERPINB1
CDH2
TCF3
MMP2
TGFB1
MSN
FN1
COL1A1
SNAI2

**Fig. S6**

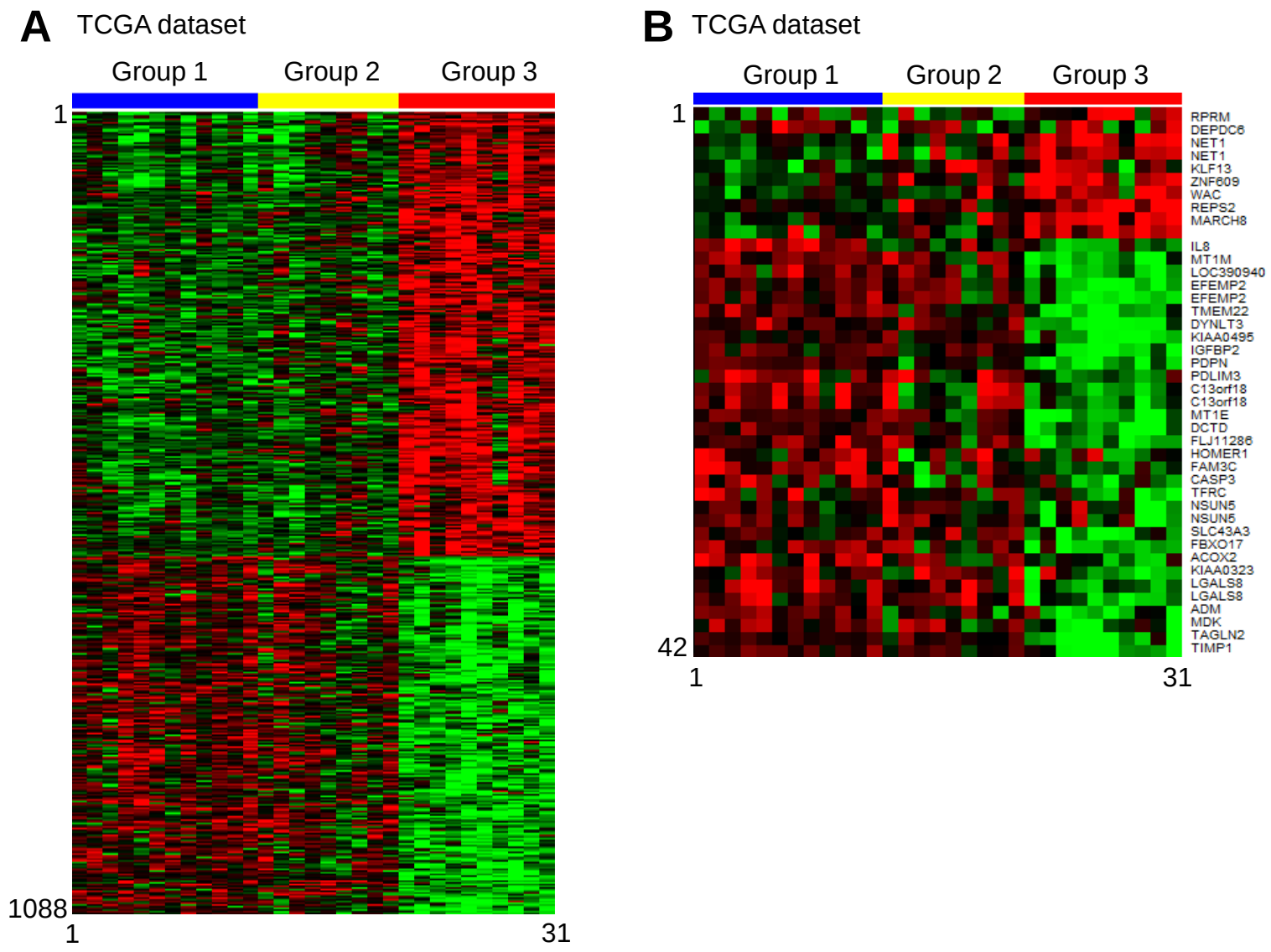

A
TCGA dataset
Group 1
Group 2
Group 3
1
1088
1
31
B
TCGA dataset
Group 1
Group 2
Group 3
1
42
1
31
RPRM
DEPDC6
NET1
NET1
KLF13
ZNF609
WAC
REPS2
MARCH8
IL8
MT1M
LOC390940
EFEMP2
EFEMP2
TMEM22
DYNLT3
KIAA0495
IGFBP2
PDPN
PDLIM3
C13orf18
C13orf18
MT1E
DCTD
FLJ11286
HOMER1
FAM3C
CASP3
TFRC
NSUN5
NSUN5
SLC43A3
FBXO17
ACOX2
KIAA0323
LGALS8
LGALS8
ADM
MDK
TAGLN2
TIMP1

**Fig. S7**

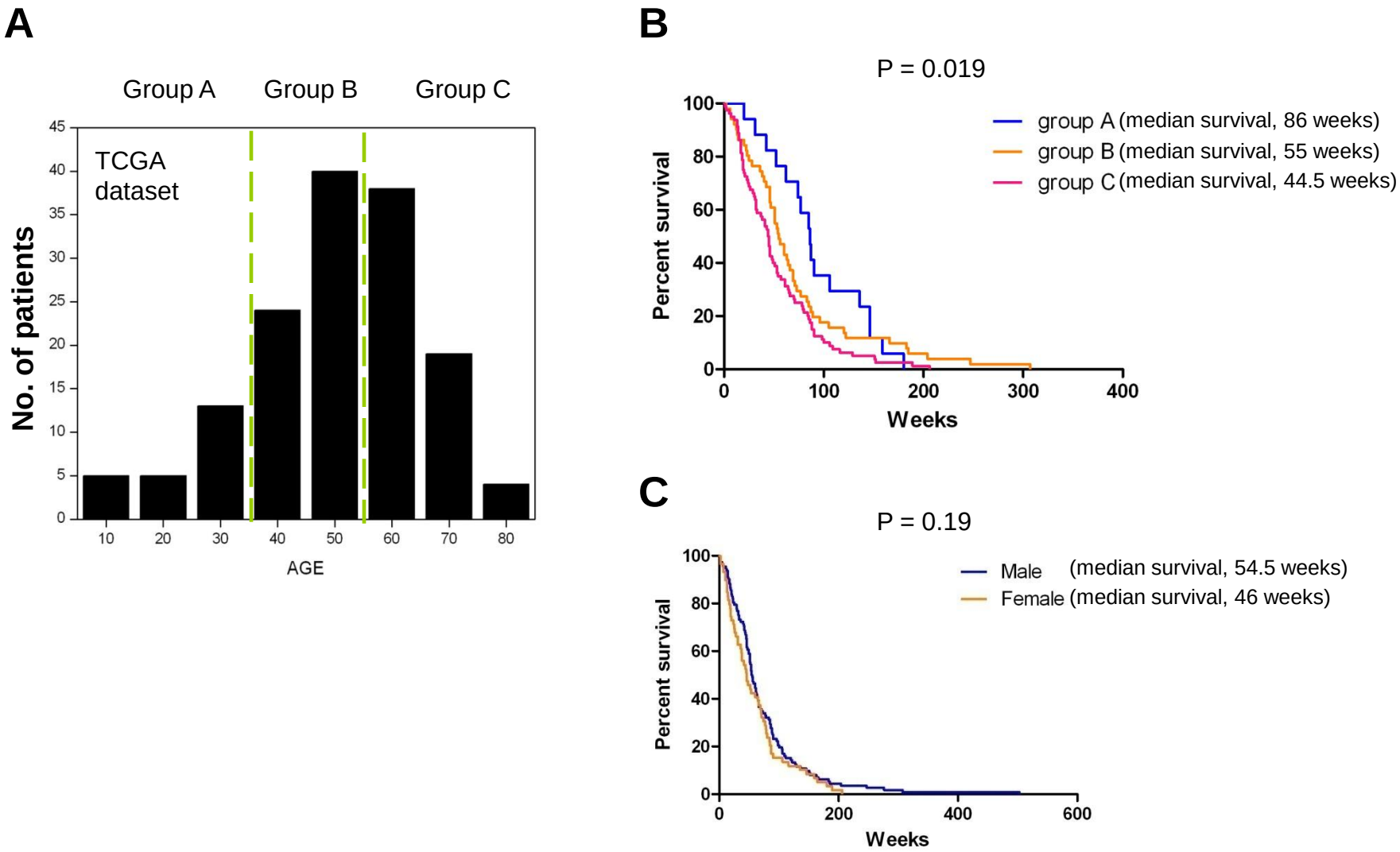

A
Group A
Group B
Group C
TCGA dataset
No. of patients
AGE
B
P = 0.019
group A (median survival, 86 weeks)
group B (median survival, 55 weeks)
group C (median survival, 44.5 weeks)
Percent survival
Weeks
C
P = 0.19
Male (median survival, 54.5 weeks)
Female (median survival, 46 weeks)
Percent survival
Weeks

**Fig. S8**

**Supplementary Figure Legends**

**Figure S1.** Hierarchical clustering analysis. **a**, Hierarchical clustering of the GSE4290 dataset of astrocytomas (II and III). Fourteen patients were classified as long-term survivors (≥ 5 years) and 14 as short-term survivors (< 5 years). The data are presented in matrix format in which rows represent individual genes and columns represent each tissue. Each cell in the matrix represents the expression level of a gene feature in an individual tissue. Red and green in cells reflect high and low expression levels, respectively. Each blue bar represents a more than 5-year survival in astrocytoma patients. **b**, Kaplan–Meier plot of overall survival of astrocytoma patients grouped on the basis of gene expression profiling. The difference between the groups was not significant ($P = 0.18$, log-rank test). **c**, Hierarchical clustering of the GSE4290 data set of oligodendrogliomas (II and III). Thirteen patients were classified as long-term survivors (≥ 5 years) and 20 as short-term survivors (< 5 years). Each blue bar represents a more than 5-year survival in oligodendroglioma patients. **d**, Kaplan–Meier plot of overall survival of oligodendrogliomas. The difference between the groups was not significant ($P = 0.99$, log-rank test).

**Figure S2.** Hierarchical clustering of astrocytomas (II and III) and oligodendrogliomas (II and III) with GBMs of GSE4290 dataset, respectively. The data are presented in matrix format in which rows represent individual genes and columns represent each tissue.

**Figure S3.** Hierarchical clustering of 77 gliomas observed in Phillips et al., 2006. a, Confirmation of hierarchical clustering of 35 signature genes as shown in Phillips et al., 2006. b, Hierarchical clustering of 77 gliomas using 41 probe sets. The data are presented in matrix format in which rows represent individual genes and columns represent each tissue.

**Figure S4.** Hierarchical clustering of 168 TCGA Core GBMs samples observed in Verhaak et al., 2010. **a**, Hierarchical clustering of 168 GBMs was used 42 probe sets. The data are presented in matrix format in which rows represent individual genes and columns represent each tissue. **b**, 168/173 samples in our set are also part of the Verhaak et al paper. We pulled the data from our data and the Verhaak dataset to make the table for the 168 samples.

**Figure S5.** Hierarchical clustering of 173 TCGA GBM tumors. Columns represent each tissue and rows represent outcomes of various prediction models as indicated. Each cell represents memberships of tissues when a particular prediction model was applied in the set. Blue represents a more than 2-year survival in GBM patients. LDA, linear discriminator analysis; SVM, support vector machines; NC, nearest centroid; NN, nearest neighbor; CCP, compound covariate prediction.

**Figure S6.** Hierarchical clustering of EMT-associated genes in TCGA dataset. EMT-associated 15 genes (29 probe sets) were generated using a two-sample t test (P = 0.05) among the three groups. Among those, the maximally expressed 15 genes were measured by average intensity across arrays. The 15 genes are presented using supervised hierarchical clustering in matrix format, where rows represent individual genes and columns represent each tissue. Each cell in the matrix represents the expression level of a gene in an individual tissue. Red and green cells reflect high and low expression levels, respectively.

**Figure S7.** Hierarchical clustering of 31 TCGA GBM tumors. a, 31 patients were classified as long-term survivors ($\geq$ 2 years). 1088 probe sets were generated using a two-sample t test (P = 0.0005) between the three groups. b, Hierarchical clustering of 31 GBMs using 42 probe sets. The data are presented in matrix format in which rows represent individual genes and columns represent each tissue.

**Figure S8.** a, Age distribution of 173 TCGA GBM patients. Patients ($n$ = 173) were classified as Group A ($\leq$ 35 years), Group B ( > 35 and $\leq$ 55 years) and Group C (> 55 years). b, Kaplan-Meier plot of overall survival in GBM patients grouped on the basis of age. The difference between the groups was significant ($P$ = 0.019, log-rank test). c, Kaplan-Meier plot of overall survival in GBM patients who were grouped on the basis of sex. The difference between the groups was not significant ($P$ = 0.19, log-rank test).