Slug-based epithelial-mesenchymal transition gene signature is associated with prolonged time to recurrence in glioblastoma.

Wei-Yi Cheng[1,7], Jessica J Kandel[2,3,4], Darrell J Yamashiro[2,4,5], Peter Canoll[5,6] and Dimitris Anastassiou[1,7§]

[1]Center for Computational Biology and Bioinformatics, Columbia University, New York, NY, USA

[2]Institute for Cancer Genetics, Columbia University, New York, NY, USA

[3]Department of Surgery, Columbia University, New York, NY, USA

[4]Department of Pediatrics, Columbia University, New York, NY, USA

[5]Department of Pathology and Cell Biology, Columbia University, New York, NY, USA

[6]Columbia Stem Cell Initiative, Columbia University, New York, NY, USA

[7]Department of Electrical Engineering, Columbia University, New York, NY 10027

[§]Corresponding author

Email addresses:

WC: wc2302@columbia.edu
JJK: jjk47@columbia.edu
DJY: dy39@columbia.edu
PC: pc561@columbia.edu
DA: da8@columbia.edu

Abstract

Background

We previously identified a precise stage-associated gene expression signature of coordinately expressed genes, including the transcription factor Slug (SNAI2) and other epithelial-mesenchymal transition (EMT) markers, present in samples from publicly available gene expression datasets in multiple cancer types. The expression levels of the co-expressed genes vary in a continuous and coordinate manner across the samples, ranging from absence of expression to strong co-expression of all genes. These data suggest that tumor cells may pass through an EMT-like process of mesenchymal transition to varying degrees.

Findings

Here we show that this signature in glioblastoma multiforme (GBM) is associated with time to recurrence following initial treatment. By analyzing data from The Cancer Genome Atlas (TCGA), we found that GBM patients who responded to therapy and had long time to recurrence had low levels of the signature in their tumor samples ($P = 3\times10^{-7}$). We also found that the signature is strongly correlated in gliomas with the putative stem cell marker CD44, and is highly enriched among the differentially expressed genes in glioblastomas vs. lower grade gliomas.

Conclusions

Our results suggest that long delay before tumor recurrence is associated with absence of the mesenchymal transition signature, raising the possibility that inhibiting this transition might improve the durability of therapy in glioma patients.

Keywords:
Glioblastoma, glioma, epithelial-mesenchymal transition, tumor recurrence, cancer stem cells

Findings

We recently identified [1] a precise multi-cancer gene expression signature, consisting of a set of genes that are coordinately overexpressed only in samples of cancer that have exceeded a particular stage, specific to each cancer type. Table 1 contains a list of the 64 genes corresponding to the top 100 probe sets (as presented in Table 4 of [1]) of the signature. The signature contains numerous epithelial-mesenchymal transition (EMT) markers [2, 3], such as the EMT-inducing transcription factor Slug (SNAI2), as well as COL5A2, FAP, POSTN, COL1A2, COL3A1, FBN1, TNFAIP6, MMP2, GREM1, BGN, CDH11, SPOCK1, DCN, COPZ2, THY1, PCOLCE, PRRX1, PDGFRB, SPARC, INHBA, COL6A2, FN1, ACTA2. However, the signature is also present even in some nonepithelial cancers, such as neuroblastoma and Ewing's sarcoma. In each dataset, the expression level of the co-expressed genes varies in a continuous manner across the samples. In a recent experiment we also confirmed that most of the genes of the signature are expressed in some xenografted human cancer cells themselves in vivo, but not in the host mouse cells. These results indicate that cancer cells can pass through a transition process to a more mesenchymal state, to varying degrees ranging from total lack of expression to strong co-expression of the genes of the signature.

The average expression level of these 64 genes can be thought of as the expression level of a metagene representing the signature. We hypothesized that this value is associated with clinical data in glioblastoma multiforme (GBM) for which there is rich such data available at The Cancer Genome Atlas (TCGA). We found that there was indeed strong association of the metagene with the phenotype "Days to Tumor Recurrence," defined as the time period from initial treatment until the date of the diagnosis or recognition of the presence and nature of the return of signs and

symptoms of cancer following a period of improvement. Patients who did not experience improvement after therapy have a "null" entry in the corresponding field.

Figure 1 shows a scatter plot in which each of the 99 samples for which the "Days to Tumor Recurrence" phenotype has a non-null entry is represented by a dot indicating the expression level of the metagene and the number of days to tumor recurrence. The figure reveals that, within the group of patients who experienced improvement after therapy, the eight patients whose tumors recurred more than three years following therapy have very low values of the expression of the metagene. Figure 2 shows a heat map of the 64 genes, where the samples are ranked in terms of the expression of the metagene and the eight patients for which time to recurrence was more than three years are highlighted in green. The rank sum for these eight patients is 1+2+6+7+9+11+16+18 = 70. This rank sum can be used as a measure of the association of the lack of expression of a gene with the "Days to Tumor Recurrence" phenotype. To evaluate its statistical significance, we randomly permuted the phenotype among the 99 samples and recalculated the rank sum, which is equivalent to finding the sum of eight randomly picked numbers from 1 to 99. This sum will be less than or equal to 70 very rarely. The estimated *P* value after doing 10 million permutations is 3×10^{-7}.

We then used the same metric (rank sum) to identify which, among the individual 64 genes of Table 1 defining the metagene have the best score, expecting that some of them would have rank sum lower than 70. Remarkably, the best scoring gene was COL5A1 with rank sum equal to 78 followed by COL6A2 with rank sum equal to 82. In other words, the score of the metagene is significantly better than that of any of its individual component genes. Even more strikingly,

after doing exhaustive search among all 12,042 genes, the top ranked gene (EFEMP2) had rank sum equal to 75, still worse than that (70) of the metagene. These results suggest that the signature identified in [1] comprises a synergistic collection of genes corresponding to a biological mechanism of mesenchymal transition, which, when absent, is associated with increased time period to tumor recurrence in GBM.

Table 2 shows a listing of the top 30 individual genes in terms of their rank sum for the "Days to Tumor Recurrence" phenotype. Nine out of these 30 genes, highlighted in Table 2, are among the 64 genes of Table 1, demonstrating the strong enrichment ($P = 3\times10^{-14}$) of EMT markers in this unbiased collection of genes associated with the phenotype.

While all cases in the TCGA dataset have been diagnosed as glioblastoma, the delayed recurrence in these eight cases is more a characteristic of lower grade gliomas. Therefore, we investigated whether lower grade gliomas are also characterized by lower levels of the signature by analyzing the NCI Repository for Molecular Brain Neoplasia Data (Rembrandt) dataset, which included gene expression from both glioblastoma as well as various types of lower grade gliomas. Table 3 demonstrates that, indeed, there is strong enrichment (seven of the 64 genes in Table 1 are among the top-ranked 30 differentially expressed genes, $P = 10^{-13}$). Furthermore, we found strong correlation between the expression levels of the metagene and the cancer stem cell marker CD44 ($P = 5\times10^{-56}$ based on fitting Pearson correlation to t-distribution). Figure 3 shows the corresponding scatter plot. Recent studies have shown that high levels of CD44 are expressed in cancer stem cells isolated from several different types of tumors [4], although this concept is still in evolution, and CD44 is also expressed in a variety of other cell types. CD44 has been

found in a cell population enriched for glioma stem cells [5]. It is also widely expressed in glioblastoma, and increased levels are associated with glioma progression and resistance to therapy [6].

Because gliomas are not epithelial cancers, and the signature has also been found in other nonepithelial cancers, such as neuroblastoma and Ewing's sarcoma, the signature represents more general biological process of mesenchymal transition.

It has recently been suggested that "stemness" in tumor cells (characterized by the ability to both self-renew as well as generate differentiated descendants) may be intimately interconnected with passing through an EMT. For example, EMT in some models was found to generate cells with properties of stem cells [7-11]. Notably, it has been shown that stem-like cells isolated from human breast cancer co-express high levels of CD44 and high levels of mesenchymal markers, including Slug [7]. Furthermore, inducing EMT in immortalized human mammary epithelial cells leads to high levels of CD44 expression in the mesenchymal-like cells [7]. Drug resistance has also been linked to the presence of cancer stem cells [9, 11-13], supporting the notion that cancer stem cells may be responsible for recurrence after therapeutic intervention. Therefore, and given the strong correlation of the mesenchymal transition signature with CD44, one possible explanation for the absence of the mesenchymal transition signature in patients with exceptionally long time to recurrence may be due to a corresponding lack of stemness in the cancer cells of these patients making it more unlikely for the cancer to recur following treatment. An alternative explanation for the observed association may be provided by the transformation towards a more mesenchymal phenotype.

The signature has been found in multiple cancers [1] and, among its component genes, Slug is the only consistently upregulated EMT-inducing transcription factor. Slug has also recently been found to be associated with invasiveness in glioma [14], consistent with the results presented here. Furthermore, when we ranked all genes in terms of their correlation (using the measure of mutual information [15]) of their expression with that of Slug in the 99 samples that we analyzed here, we found that, remarkably, the top eight entries (COL6A3, COL3A1, LUM, COL5A1, COL1A2, COL6A2, COL1A1, PCOLCE) were all genes included in both Tables 1 as well as Table 2, further supporting the hypothesis that Slug might be a master regulator of the biological mechanism responsible for the signature.

The same signature was also found to be predictive of neoadjuvant therapy in breast cancer - see, e.g. additional file 6 of [1], in which 7 of 8 samples in the cluster on the left side of the heat map (with low levels of the signature) had good response to therapy, while 12 out of 14 samples in the second cluster (with high levels of the signature) were resistant.

Analysis of gene expression data has resulted in classification into various subtypes of glioblastomas [16, 17], also present in lower grade gliomas [18], with distinct features, each of which is characterized by the presence of particular genes. Interestingly, CD44 was found enriched either in the mesenchymal subtypes or in glioblastomas in all these cases. The feature of our current results, however, is that the mesenchymal transition signature used in this paper reflects a biological process applicable to multiple cancer types, as it was derived by analyzing its presence in many different cancers [1], as opposed to using classification methods on glioma

samples alone to identify subtypes. Furthermore, the association with the phenotype is found in the absence, rather than the presence, of the signature.

The observations that (a) all GBM patients with exceptionally long time to recurrence had extremely low levels of the mesenchymal transition gene signature, and (b) the mesenchymal transition signature is strongly enriched among the genes underexpressed in lower grade gliomas as compared to glioblastomas, suggest that targeting the underlying biological mechanism might supply a novel approach for adjuvant treatment of gliomas. Further, the ability to precisely identify components of the gene signature provides unique opportunities for identifying potential targets for such treatment.

References

1. Kim H, Watkinson J, Varadan V, Anastassiou D: Multi-cancer computational analysis reveals invasion-associated variant of desmoplastic reaction involving INHBA, THBS2 and COL11A1. *BMC Med Genomics* 2010, 3:51.
2. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, Hartwell K, Onder TT, Gupta PB, Evans KW, et al: Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc Natl Acad Sci U S A* 2010, 107:15449-15454.
3. Jechlinger M, Grunert S, Tamir IH, Janda E, Ludemann S, Waerner T, Seither P, Weith A, Beug H, Kraut N: Expression profiling of epithelial plasticity in tumor progression. *Oncogene* 2003, 22:7155-7169.
4. Zoller M: CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? *Nat Rev Cancer* 2011, 11:254-267.
5. Anido J, Saez-Borderias A, Gonzalez-Junca A, Rodon L, Folch G, Carmona MA, Prieto-Sanchez RM, Barba I, Martinez-Saez E, Prudkin L, et al: TGF-beta Receptor Inhibitors Target the CD44(high)/Id1(high) Glioma-Initiating Cell Population in Human Glioblastoma. *Cancer Cell* 2010, 18:655-668.
6. Xu Y, Stamenkovic I, Yu Q: CD44 attenuates activation of the hippo signaling pathway and is a prime therapeutic target for glioblastoma. *Cancer Res* 2010, 70:2455-2464.
7. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, et al: The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 2008, 133:704-715.
8. Morel AP, Lievre M, Thomas C, Hinkal G, Ansieau S, Puisieux A: Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One* 2008, 3:e2888.
9. Singh A, Settleman J: EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene* 2010, 29:4741-4751.
10. Scheel C, Weinberg RA: Phenotypic plasticity and epithelial-mesenchymal transitions in cancer and normal stem cells? *Int J Cancer* 2011, 129:2310-2314.
11. Alison MR, Lim SM, Nicholson LJ: Cancer stem cells: problems for therapy? *J Pathol* 2011, 223:147-161.
12. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A, Rimm DL, Wong H, Rodriguez A, Herschkowitz JI, et al: Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc Natl Acad Sci U S A* 2009, 106:13820-13825.
13. Buck E, Eyzaguirre A, Barr S, Thompson S, Sennello R, Young D, Iwata KK, Gibson NW, Cagnoni P, Haley JD: Loss of homotypic cell adhesion by epithelial-mesenchymal transition or mutation limits sensitivity to epidermal growth factor receptor inhibition. *Mol Cancer Ther* 2007, 6:532-541.
14. Yang HW, Menon LG, Black PM, Carroll RS, Johnson MD: SNAI2/Slug promotes growth and invasion in human gliomas. *BMC Cancer* 2010, 10:301.
15. Cover TM, Thomas JA: *Elements of information theory.* 2nd edn. Hoboken, N.J.: Wiley-Interscience; 2006.
16. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, et al: Integrated genomic analysis identifies clinically relevant subtypes of

glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 2010, 17:98-110.

17. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, Misra A, Nigro JM, Colman H, Soroceanu L, et al: Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 2006, 9:157-173.

18. Cooper LA, Gutman DA, Long Q, Johnson BA, Cholleti SR, Kurc T, Saltz JH, Brat DJ, Moreno CS: The proneural molecular signature is enriched in oligodendrogliomas and predicts improved survival among diffuse gliomas. *PLoS One* 2010, 5:e12548.

Figure captions

Figure 1. Scatter plot for Days to Tumor Recurrence vs. expression of the mesenchymal transition metagene

Each dot in the scatter plot represents one of the 99 patients for which the "Days to Tumor Recurrence" phenotype has a non-null entry. The horizontal axis measures the average of the RMA-normalized expression levels of the 64 genes shown in Table 1. The vertical axis measures the days to tumor recurrence and the horizontal dotted line is drawn at the 3 year cutoff point.

Figure 2. Heat map of the components of the mesenchymal transition metagene in glioblastoma

The 99 samples are ranked in terms of the average expression level of the genes shown in Table 1. The eight patients for which time to recurrence was more than three years are highlighted in green at the 1^{st}, 2^{nd}, 6^{th}, 7^{th}, 9^{th}, 11^{th}, 16^{th}, and 18^{th} position, resulting in the rank sum of 70.

Figure 3. Scatter plot for the expression levels of CD44 vs. the mesenchymal transition metagene

Each dot in the scatter plot represents a glioma sample from the NCI Repository for Molecular Brain Neoplasia Data (Rembrandt) dataset. Dots are color coded red for glioblastomas and blue for lower grade gliomas. Expression levels are RNA normalized.

Figure 1. Scatter plot for Days to Tumor Recurrence vs. expression of the mesenchymal transition metagene

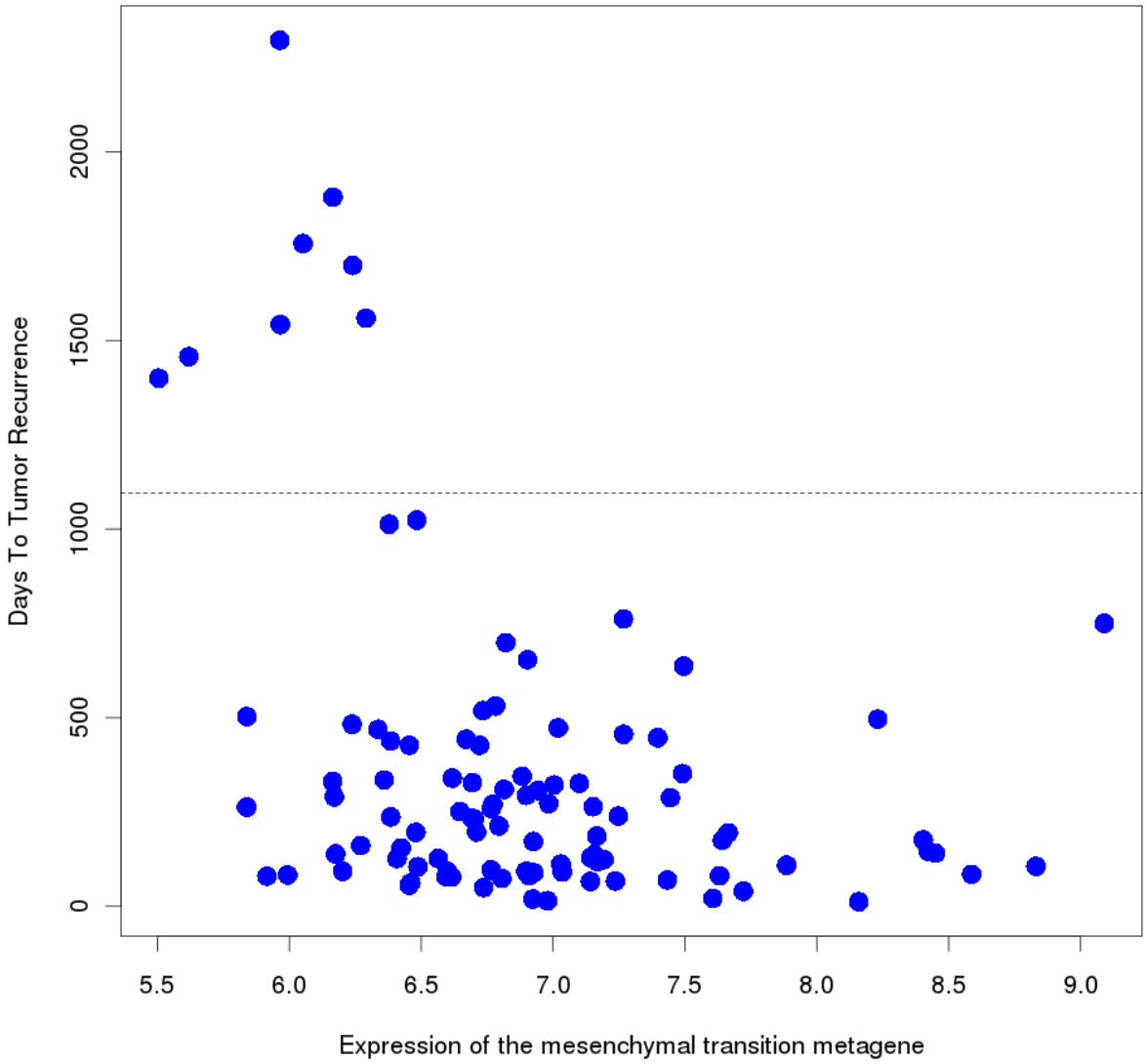

Figure 2. Heat map of the components of the mesenchymal transition metagene in glioblastoma

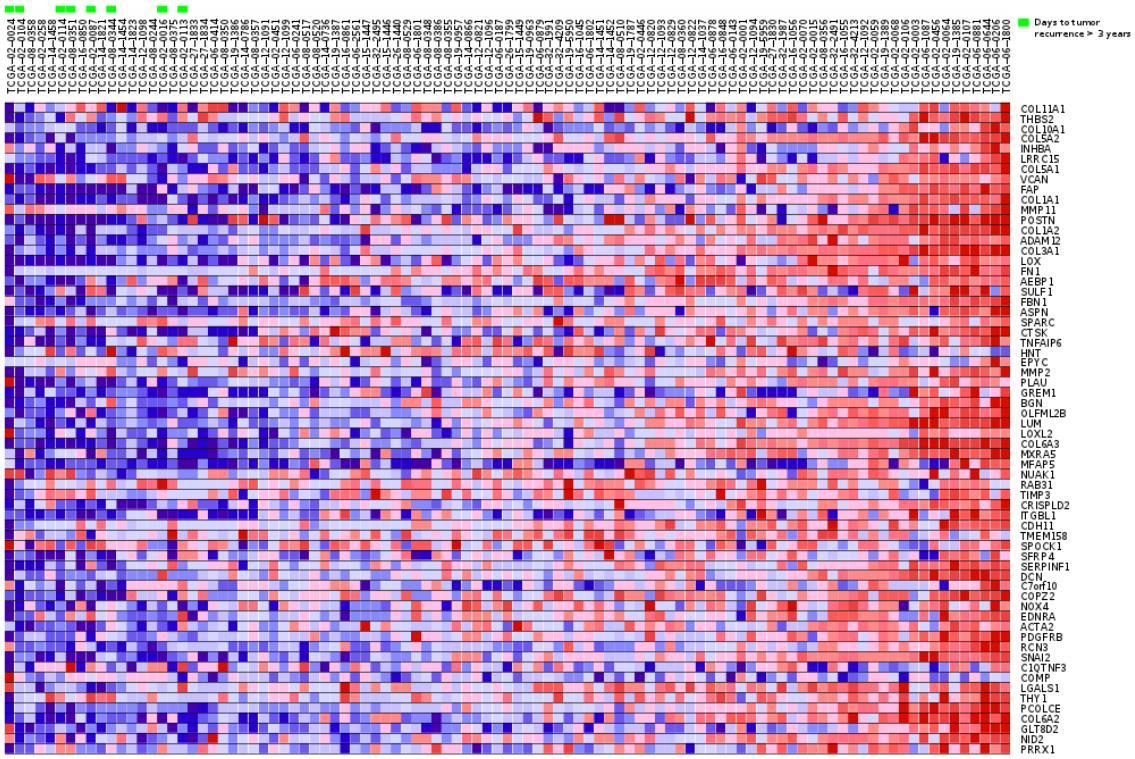

Figure 3. Scatter plot for the expression levels of CD44 vs. the mesenchymal transition metagene

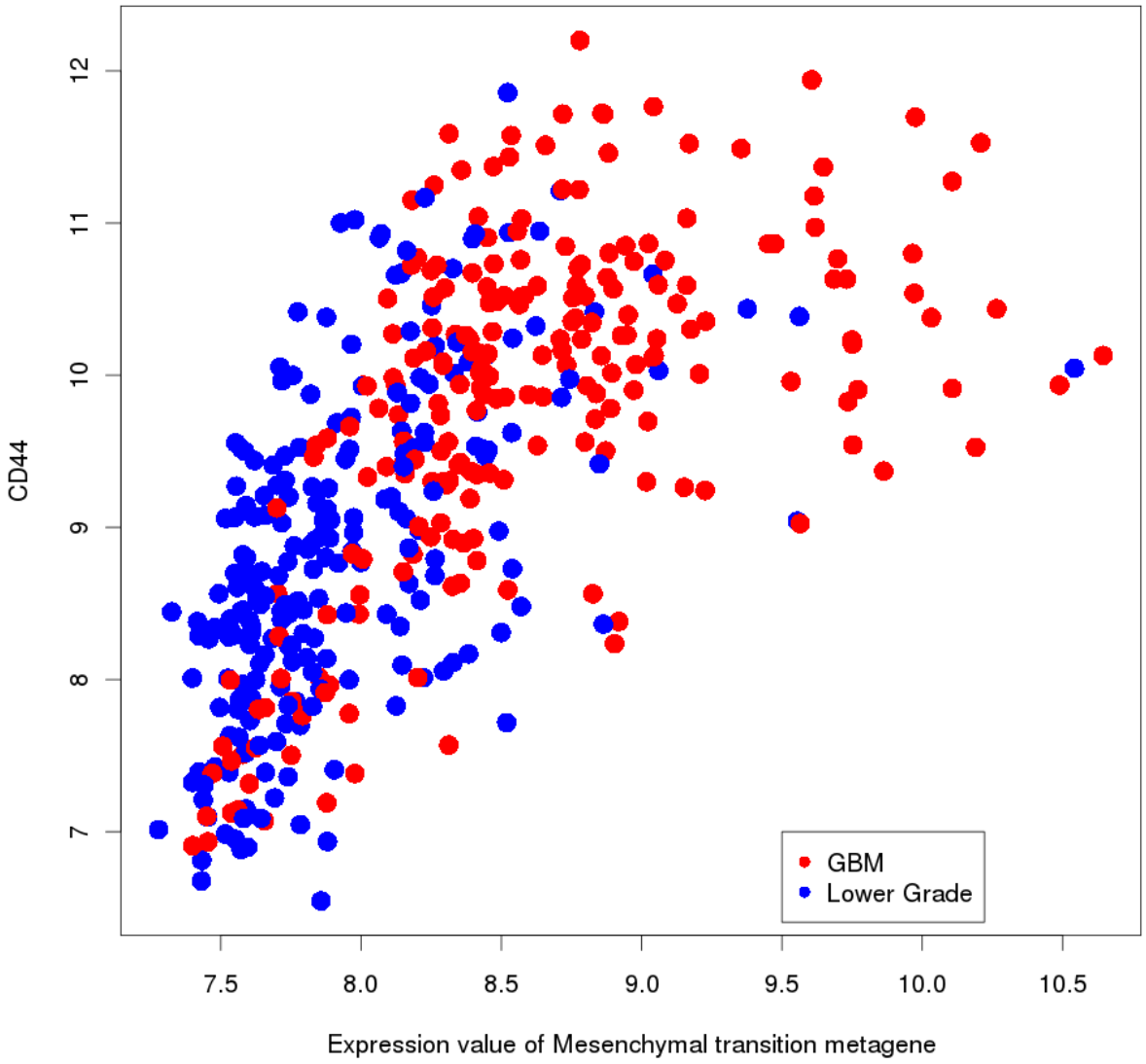

Table 1. Genes comprising the Slug-based EMT signature.

Rank	Gene	Rank	Gene
1	COL11A1	33	LOXL2
2	THBS2	34	COL6A3
3	COL10A1	35	MXRA5
4	COL5A2	36	MFAP5
5	INHBA	37	NUAK1
6	LRRC15	38	RAB31
7	COL5A1	39	TIMP3
8	VCAN	40	CRISPLD2
9	FAP	41	ITGBL1
10	COL1A1	42	CDH11
11	MMP11	43	TMEM158
12	POSTN	44	SPOCK1
13	COL1A2	45	SFRP4
14	ADAM12	46	SERPINF1
15	COL3A1	47	DCN
16	LOX	48	C7orf10
17	FN1	49	COPZ2
18	AEBP1	50	NOX4
19	SULF1	51	EDNRA
20	FBN1	52	ACTA2
21	ASPN	53	PDGFRB
22	SPARC	54	RCN3
23	CTSK	55	SNAI2
24	TNFAIP6	56	C1QTNF3
25	HNT	57	COMP
26	EPYC	58	LGALS1
27	MMP2	59	THY1
28	PLAU	60	PCOLCE
29	GREM1	61	COL6A2
30	BGN	62	GLT8D2
31	OLFML2B	63	NID2
32	LUM	64	PRRX1

Table 2. Top genes in terms of the rank sum for the "Days to Tumor Recurrence" phenotype

EFEMP2	75
CD248	78
COL5A1	78
IL7R	78
MYH9	81
COL6A2	82
FLNC	82
AKAP12	87
TREM1	87
COL1A2	88
PSCDBP	88
S100A8	91
CLEC2B	92
GLIPR1	98
COL6A3	99
THBD	99
CALD1	102
CD163	102
EFEMP1	102
ENPEP	103
PCOLCE	103
TMEM5	103
SDCBP	104
COL1A1	105
LUM	105
TNC	106
VNN1	106
CARS	107
FN1	107
COL3A1	108

Table 3. Top differentially expressed gene in glioblastomas vs. lower grade gliomas

Probe Set ID	Gene Symbol	Fold Change
209395_at	CHI3L1	10.36
202718_at	IGFBP2	6.07
210809_s_at	**POSTN**	5.77
201666_at	TIMP1	5.70
1556499_s_at	**COL1A1**	5.69
215076_s_at	**COL3A1**	5.59
202404_s_at	**COL1A2**	5.17
206157_at	PTX3	4.90
201012_at	ANXA1	4.87
202237_at	NNMT	4.82
211527_x_at	VEGFA	4.73
221898_at	PDPN	4.65
202912_at	ADM	4.65
215446_s_at	**LOX**	4.43
202345_s_at	FABP5	4.41
226517_at	BCAT1	4.30
203729_at	EMP3	4.14
202018_s_at	LTF	4.05
227697_at	SOCS3	3.96
211981_at	COL4A1	3.64
209156_s_at	**COL6A2**	3.62
201505_at	LAMB1	3.59
226237_at	226237_at	3.59
236028_at	IBSP	3.57
201744_s_at	**LUM**	3.53
225681_at	CTHRC1	3.52
203645_s_at	CD163	3.51
211964_at	COL4A2	3.49
201110_s_at	THBS1	3.44
208949_s_at	LGALS3	3.42