

# Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data

Lawrence P. Petalidis,<sup>1</sup> Anastasis Oulas,<sup>2,3</sup> Magnus Backlund,<sup>6</sup> Matthew T. Wayland,<sup>4</sup> Lu Liu,<sup>1</sup> Karen Plant,<sup>1</sup> Lisa Happerfield,<sup>1</sup> Tom C. Freeman,<sup>5</sup> Panayiota Poirazi,<sup>2</sup> and V. Peter Collins<sup>1</sup>

<sup>1</sup>Division of Molecular Histopathology, Department of Pathology, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom; <sup>2</sup>Institute for Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas; <sup>3</sup>Graduate Program in Molecular Biology and Biomedicine, Department of Biology, University of Crete, Heraklion, Crete, Greece; <sup>4</sup>Cambridge Centre for Neuropsychiatric Research, University of Cambridge, Cambridge, United Kingdom; <sup>5</sup>Division of Pathway Medicine, College of Medicine, University of Edinburgh, Edinburgh, Scotland, United Kingdom; and <sup>6</sup>Department of Oncology-Pathology, Karolinska Hospital, Karolinska Institute, Stockholm, Sweden

## Abstract

Histopathologic grading of astrocytic tumors based on current WHO criteria offers a valuable but simplified representation of oncologic reality and is often insufficient to predict clinical outcome. In this study, we report a new astrocytic tumor microarray gene expression data set ( $n = 65$ ). We have used a simple artificial neural network algorithm to address grading of human astrocytic tumors, derive specific transcriptional signatures from histopathologic subtypes of astrocytic tumors, and assess whether these molecular signatures define survival prognostic subclasses. Fifty-nine classifier genes were identified and found to fall within three distinct functional classes, that is, angiogenesis, cell differentiation, and lower-grade astrocytic tumor discrimination. These gene classes were found to characterize three molecular tumor subtypes denoted **ANGIO**, **INTER**, and **LOWER**. Grading of samples

using these subtypes agreed with prior histopathologic grading for both our data set (96.15%) and an independent data set. Six tumors were particularly challenging to diagnose histopathologically. We present an artificial neural network grading for these samples and offer an evidence-based interpretation of grading results using clinical metadata to substantiate findings. The prognostic value of the three identified tumor subtypes was found to outperform histopathologic grading as well as tumor subtypes reported in other studies, indicating a high survival prognostic potential for the 59 gene classifiers. Finally, 11 gene classifiers that differentiate between primary and secondary glioblastomas were also identified. [Mol Cancer Ther 2008;7(5):1013–24]

## Introduction

Astrocytic tumors of malignancy grades 2 to 4 are collectively termed diffusely infiltrating astrocytomas and include diffuse astrocytoma (malignancy grade 2; "A"), anaplastic astrocytoma (malignancy grade 3; "AA"), and glioblastoma (malignancy grade 4; "GB"). A total of four malignancy grades are recognized by the WHO system, with grades 1 and 4 tumors being the biologically least and most aggressive tumor grades, respectively (1, 2). Glioblastoma commonly occurs *de novo* (also called primary glioblastoma) but may also result from the progression of lower-grade tumors to higher malignancy grades. Glioblastoma shows the greatest range of genetic abnormalities, with common changes in the *de novo* tumors, including homozygous deletion of *CDKN2A*, *CDKN2B*, and *p14<sup>ARF</sup>* (9p21), loss of one allele and mutation of the retained allele of *PTEN* (10q23), and amplification of the *EGFR* gene (7p12; ref. 2).

Use of expression microarray data in brain tumor classification/clustering (3) and survival prognosis (4–6) has received significant interest in the last few years. Approaches include statistical methods for gene set identification and tumor classification (7); principal component analysis and *t* test for the selection of differentially expressed genes involved in astrocytoma progression (8); *k* means along with multidimensional scaling for discriminating between glioblastomas, lower-grade astrocytomas, and other glioma types, such as oligodendrogliomas (9); hierarchical clustering (3, 4, 9, 10); *k* nearest neighbor for classification of high-grade gliomas and outcome prognosis (5); gene voting for survival prediction of the diffusely infiltrating gliomas (4); and others. Expression profiling has identified molecular as well as genetic subtypes associated with tumor grade, progression, and patient survival (8, 10). Although astrocytic tumors continue to be defined by histologic criteria, reports that expression profiles predict survival better than histologic grade (4, 5, 11) provide

Received 3/13/07; revised 1/11/08; accepted 3/26/08.

**Grant support:** Cancer Research UK, UK Medical Research Council, The Jacqueline Seroussi Memorial Foundation for Cancer Research, Samantha Dickson Research Trust, Ludwig Institute for Cancer Research, General Secretariat for Research and Technology, Hellas (project PENED 03ED842), and EMBO Young Investigator Program.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**Note:** L. P. Petalidis, A. Oulas, P. Poirazi, and V.P. Collins contributed equally to this work.

**Requests for reprints:** Panayiota Poirazi, Institute for Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, Vassilikia Vouton, P.O. Box 1385, GR 711 10 Heraklion, Crete, Greece. E-mail: poirazi@imbb.forth.gr

Copyright © 2008 American Association for Cancer Research.

doi:10.1158/1535-7163.MCT-07-0177

support for the hypothesis that tumors defined morphologically represent a mix of molecular genetic subtypes. Most of these studies, however, have compared diffusely infiltrating astrocytomas with tumors of mixed or non-astrocytic origin, have not included lower-grade (2) tumors (4, 5, 7, 11), or have limited their efforts to a single tumor grade (3, 6). Several of these studies have also compared tumor tissue to normal brain, a task of arguable relevance when taking into account the vast differences in cellular composition between the two tissues. Moreover, studies have often focused on questions related more to the use of expression data toward general brain tumor classification rather than malignancy grading of diffusely infiltrating astrocytic tumors per se. Finally, discordances between histopathology and expression-based tumor classification for a given tumor set have seldom been interpreted or substantiated with thorough clinical and/or molecular evidence.

Using a new gene expression data set originating from 65 highly annotated tumors and a simple artificial neural network (ANN) algorithm in the form of a single-layer perceptron, we address grading of human astrocytic tumors, derive specific transcriptional signatures from histopathologic subtypes of astrocytic tumors, and assess whether these molecular signatures define survival prognostic subclasses. We validate our approach with several independent data sets and offer valuable insight into the tumor biology and gene expression-based grading of astrocytomas.

## Materials and Methods

### Tumor Samples, RNA Isolation, and Hybridization to Affymetrix U133A GeneChips

The tumor set consisted of 2 pilocytic astrocytomas (WHO grade 1 "PA"), 5 diffuse astrocytomas (WHO grade 2 "A"), 15 anaplastic astrocytomas (WHO grade 3 "AA"), and 39 glioblastomas (WHO grade 4 "GB"). This sample distribution reflects tissue availability and relative frequency of diagnosis per tumor grade. Four additional samples graded as AA that were exceptionally challenging to grade by histopathology were treated as separate "problem" cases. Histopathologic diagnoses were made according to WHO criteria (1) by V.P.C. RNA from the 65 human astrocytic tumor samples was extracted using guanidine isothiocyanate ultracentrifugation as described previously (12). RNA quality was assessed using an Agilent Bioanalyzer 2100 (Agilent Technologies). For each tumor sample, 7  $\mu$ g RNA was used to generate double-stranded cDNA, which was subsequently *in vitro* transcribed to produce biotin-labeled cRNA using the ENZO BioArray HighYield kit. cRNA (15  $\mu$ g) was fragmented and hybridized to Affymetrix HG-U133A GeneChips (Affymetrix). GeneChips were washed, stained, and scanned as described in the manufacturer's manual. Quality of prefragmentation and postfragmentation cRNA was assessed using an Agilent Bioanalyzer 2100 (Agilent Technologies).

### Expression Microarray Data Analysis

Raw data (CEL files) were imported into "R," a freely available environment for statistical computing (13). Nor-

malization and computation of expression measures was done using the justRMA function within the Affy package of Bioconductor (14). All expression data have been submitted to GEO (15) in a MIAME-compliant fashion (accession no. GSE1993). Annotation of probe set lists was done using EASE (16).

### Validation of Results Using Quantitative PCR

Quantitative PCR (QPCR) was done on a LightCycler machine (Roche) using DNA Master SYBR Green I (Roche Molecular Biochemicals or Sigma) according to the manufacturer's protocol. Primers were ordered from MWG. Double-stranded cDNA used as a template was the same as that used for cRNA target preparation. One microliter of this cDNA was diluted 1:200 for generation of the final template used. Validation was done on a subset of 23 tumors (15 GB and 8 AA) that were part of the original tumor data set assessed. Assays were done in duplicate. The raw data produced by QPCR referred to the number of cycles required for reactions to reach exponential phase as determined using the RelQuant software (Roche). Expression of *MYO1C* was used for normalization of the QPCR data. Mean expression fold change differences between tumor groups were calculated using the  $2^{-\Delta\Delta CT}$  method (17). Primer sequences: *phosphoprotein enriched in astrocytes 15 (PEA15)* 5'-GAGCAGCCAGCGTTAGATGC-3' and 3'-GGAGGTGTTTCAAGACCAGGG-5' and *adrenomedullin (ADM)* 5'-GCAGAAGAATCCGAGTGTTC-3' and 3'-AATCAGTTTGTGGGCGAGCACG-5'.

### Tissue Array Generation and Immunohistochemistry

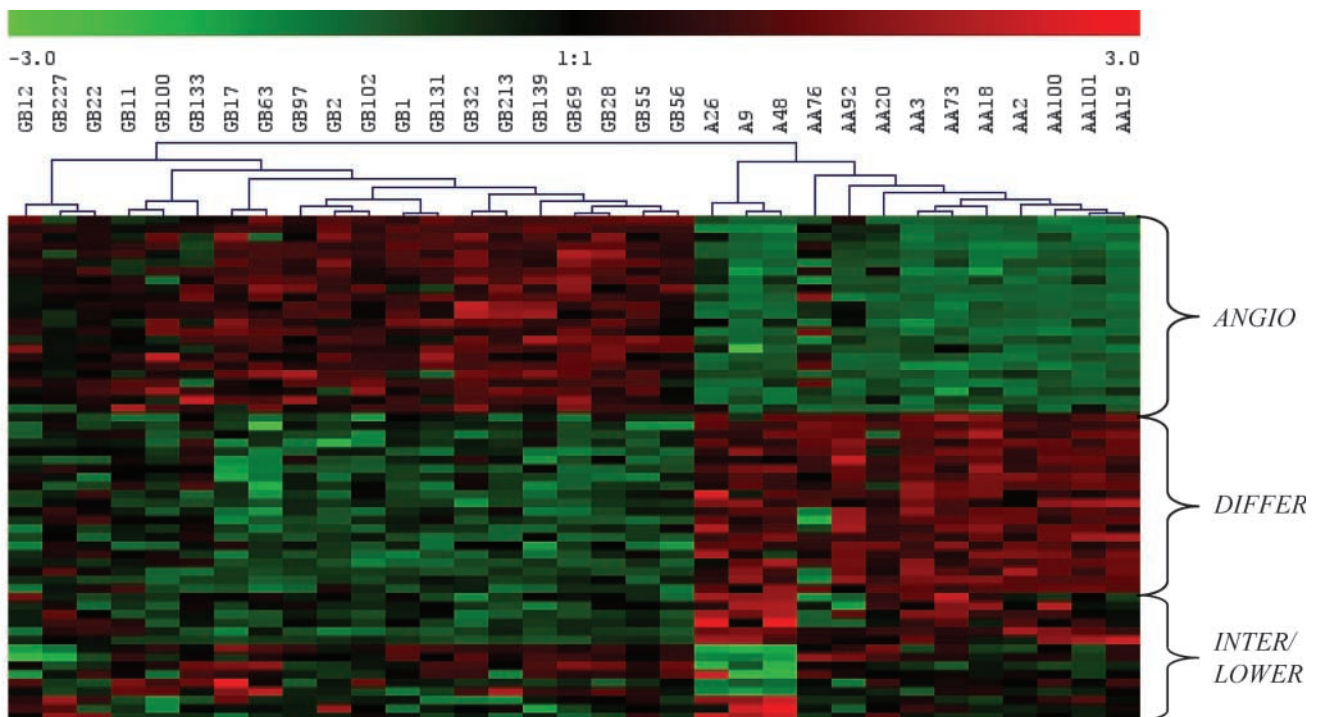
Cores ( $n = 2$  for 57 tumors from our data set) of 0.6 mm diameter were taken from paraffin-embedded tumor tissue and arrayed into a fresh paraffin block using a manual tissue arrayer (Beecher Instruments). Areas identified on H&E-stained sections to have high tumor cell content were used. Ten nonneoplastic tissue cores with minimal or no tumor cell content were also included. Immunohistochemistry for *ADM* (1:50; Abcam) and *PEA15* (1:500) was done as described previously (18, 19).

### ANN Model and Statistical Analysis

A single-layer perceptron was used for grading the tumor tissue samples. The number of inputs was equal to the number of classifier genes and the output layer consisted of a single neuron with a sigmoidal activation function. Initial weight values were chosen randomly and training was done using a standard gradient descent learning rule (or Delta rule) with learning rate  $\eta = 0.05$ . Calibration was done via leave-one-out cross-validation. The weight values were updated after every sample and the calibration was terminated after 100 passes (epochs) through the entire training set. The resulting variables for a completed training define a "model" (also see Supplementary Data).<sup>7,8</sup>

<sup>7</sup> Supplementary material for this article is available at Molecular Cancer Therapeutics Online (<http://mct.aacrjournals.org/>).

<sup>8</sup> The source code for the ANN and visualization methods is available from <http://www.imbb.forth.gr/people/poirazi/software.html>.



**Figure 1.** Hierarchical clustering of 33 training samples (20GB, 10 AA, and 3 A) using 59 probe sets selected by S2N. MeV (33) was used to perform hierarchical clustering using Euclidean distance and complete linkage algorithm. Samples are labeled with their respective grades and genes are labeled according to the molecular tumor subtype(s) (*ANGIO*, *DIFFER*, or *INTER/LOWER*) that they characterize. Gene expression values are standardized to a mean of 0 and SD of 1. *Red*, higher expression relative to *green*.

The Kaplan-Meier method was used to estimate the survival distributions (20). Log-rank tests were used to test the difference between survival groups. For all of the analyses, a  $P < 5.0e^{-2}$  was accepted as significant. Statistical analyses were carried out with the freely available software package R.

## Results

### Selection of Classifier Genes

**Training the ANN to Distinguish between Different Astrocytic Tumor Grades and Concurrent Selection of Classifier Genes.** To train the neural network, tumor samples were randomly split into two sets in a way that approximately preserves the sample distribution across each tumor grade. The first 20 GB, 10 AA, and 3 A were used as a training set and 19 GB, 5 AA, and 2 A were used as a test set. A further test group of 6 astrocytic tumors comprised 4 AA that had proved difficult to grade histopathologically and two samples belonging to grade 1 PA.

Training/calibration was done in an all-pairs approach whereby the single problem of learning to differentiate between three grades (GB, AA, and A) was narrowed down to multiple two-grade problems (Supplementary Fig. SA).<sup>7</sup> The 33 training samples were split into three sample groups each comprising two tumor grades, that is,

(a) GB-AA, (b) AA-A, and (c) GB-A. Three different types of ANN models (A, B, and C) were then trained, each corresponding to their respective sample groups. For each of these model types, genes that showed differential expression between the two grades in question were selected using the signal-to-noise method (21) on the entire U133A chip genome. Training performance and optimum number of genes required for grading were evaluated using leave-one-out cross-validation. For every leave-one-out run, genes were ranked according to signal-to-noise (taken over all but the left out sample) and then the grading success rate was determined using increasing numbers of these ranked genes. Leave-one-out cross-validation success rates optimized to 93.3%, 84.6%, and 95.6% using a total of 44, 9, and 7 probe sets for the GB-AA, AA-A and GB-A grade comparisons, respectively (see leave-one-out plots; Supplementary Fig. SB).<sup>7</sup>

Pooling of all probe sets and elimination of redundancies resulted in a total of 59 unique probe sets. As anticipated, hierarchical clustering of all training samples using the above probe sets revealed clear distinctions between the GB, AA, and A tumor grades and further defined three functional gene classes that delineate three molecular tumor subtypes (Fig. 1; see next section for details). The trained/calibrated ANN models (see Materials and Methods) were subsequently used for grading of the test set.

Table 1. Three sets of selected genes each derived from one of the three pairwise tumor grade comparisons

Gene symbol	Gene name	Bayesian <i>P</i>	Mean expression fold change	Gene class
(A) GB-AA (ANGIO/DIFFER genes)				
ADM	Adrenomedullin	2.62E-05	11.79	ANGIO
TIMP1	Tissue inhibitor of metalloproteinase 1	1.51E-08	11.56	ANGIO
FABP5	Fatty acid binding protein 5	1.85E-04	9.41	ANGIO
EMP3	Epithelial membrane protein 3	5.27E-07	7.58	ANGIO
PDPN	Podoplanin	3.22E-05	6.00	ANGIO
LGALS3	Lectin galactoside binding soluble 3 (galectin 3)	1.02E-05	5.86	ANGIO
LGALS1	Lectin galactoside binding soluble 1 (galectin 1)	1.02E-05	4.41	ANGIO
PDGFA	Platelet-derived growth factor $\alpha$ polypeptide	2.03E-05	4.09	ANGIO
PLAT	Plasminogen activator tissue	6.60E-05	3.97	ANGIO
EFEMP2	EGF-containing fibulin-like extracellular matrix protein 2	2.20E-06	3.92	ANGIO
COL5A2	Collagen type V $\alpha$ 2	3.71E-05	3.72	ANGIO
COL5A2	Collagen type V $\alpha$ 2	1.02E-05	3.60	ANGIO
DDA3	Differential display and activated by p53	2.06E-05	3.53	ANGIO
TAGLN2	Transgelin 2	3.15E-05	3.19	ANGIO
DUSP6	Dual-specificity phosphatase 6	5.49E-05	3.14	ANGIO
LDHA	Lactate dehydrogenase A	7.84E-05	2.84	ANGIO
PLP2	Proteolipid protein 2	6.34E-05	2.74	ANGIO
EFEMP2	EGF-containing fibulin-like extracellular matrix protein 2	7.34E-05	2.43	ANGIO
CENTD3	Centaurin $\delta$ 3	2.57E-04	2.42	ANGIO
KIAA0495	KIAA0495	1.69E-04	2.15	ANGIO
DAG1	Dystroglycan 1 (dystrophin-associated glycoprotein 1)	2.73E-05	1.80	ANGIO
ZYX	Zyxin	1.46E-04	1.78	ANGIO
OSBPL10	Oxysterol binding protein-like 10	2.63E-04	1.75	ANGIO
CUTC	cutC copper transporter homologue	1.74E-05	-1.59	DIFFER
TNKS2	TRF1-interacting ankyrin-related ADP-ribose polymerase 2	6.60E-05	-1.67	DIFFER
HSA9761	Dimethyladenosine transferase	8.27E-05	-1.71	DIFFER
KIAA1279	KIAA1279	2.03E-05	-1.76	DIFFER
RPL22	Ribosomal protein L22	2.83E-04	-1.81	DIFFER
ENAH	Enabled homologue	6.26E-05	-1.82	DIFFER
ZMYND11	Zinc finger MYND domain containing 11	4.53E-05	-1.87	DIFFER
HNRPH3	Heterogeneous nuclear ribonucleoprotein H3	3.65E-05	-1.88	DIFFER
RPL22	Ribosomal protein L22	2.62E-05	-1.93	DIFFER
CLASP2	Cytoplasmic linker-associated protein 2	2.81E-04	-2.05	DIFFER
USH1C	Usher syndrome 1c (autosomal recessive severe)	1.02E-05	-2.10	DIFFER
RAP2A	RAP2A	7.90E-04	-2.10	DIFFER
ALCAM	Activated leukocyte cell adhesion molecule	4.72E-03	-2.14	DIFFER
ABII	Abl-interactor 1	8.27E-05	-2.16	DIFFER
PAR3	par-3 partitioning defective 3 homologue	3.43E-06	-2.30	DIFFER
CRYAB	Crystallin $\alpha$ B	5.80E-05	-2.72	DIFFER
NAP1L3	Nucleosome assembly protein 1-like 3	1.77E-04	-2.88	DIFFER
NET1	Neuroepithelial cell transforming gene 1	2.20E-06	-2.93	DIFFER
C20ORF42	Chromosome 20 open reading frame 42	2.87E-04	-3.09	DIFFER
BMP2	Bone morphogenetic protein 2	5.49E-05	-3.27	DIFFER
ADCY2	Adenylate cyclase 2 (brain)	3.05E-05	-3.79	DIFFER
(B) GB-A (INTER/LOWER genes)				
SLC34A1	Solute carrier family 34 (sodium phosphate) member 1	1.46E-02	-1.35	LOWER
RSNL2	Restin-like 2	1.88E-02	-1.39	LOWER
REPS2	RALBP1-associated EPS domain containing 2	1.53E-03	-1.94	LOWER
SLCO1A2	Solute carrier organic anion transporter family member 1A2	4.46E-03	-2.06	LOWER
PEA15	Phosphoprotein enriched in astrocytes 15	1.93E-03	-2.12	LOWER
USH1C	Usher syndrome 1C	1.53E-03	-3.49	LOWER
(C) AA-A (INTER/LOWER genes)				
B2M	$\beta$ 2-microglobulin	3.69E-01	2.50	INTER
SCP2	Sterol carrier protein 2	4.79E-01	1.81	INTER

(Continued on the following page)

**Table 1. Three sets of selected genes each derived from one of the three pairwise tumor grade comparisons (Cont'd)**

Gene symbol	Gene name	Bayesian <i>P</i>	Mean expression fold change	Gene class
(C) AA-A (INTER/LOWER genes)				
<i>DDOST</i>	<i>Dolichyl-diphosphooligosaccharide-protein glycosyltransferase</i>	5.68E-01	1.79	INTER
<i>NPTN</i>	<i>Neuroplastin</i>	8.49E-01	1.39	INTER
<i>TAP2</i>	<i>Transporter 2 ATP-binding cassette subfamily B</i>	6.92E-01	1.39	INTER
<i>DNAJA3</i>	<i>DnaJ (hsp40) homologue subfamily A, member 3</i>	6.48E-01	1.25	INTER
<i>ANKS1b</i>	<i>Ankyrin repeat and sterile <math>\alpha</math> motif domain containing 1b</i>	5.17E-01	-1.21	LOWER
—	<i>DKFZp434M083</i>	5.68E-01	-1.22	LOWER
<i>ANK3</i>	<i>Ankyrin 3 node of Ranvier</i>	9.07E-02	-1.93	LOWER

### Expression Profiles of Gene Classifiers Selected during Training Define Three Molecular Tumor Subtypes.

A thorough examination of the selected gene classifiers, most of which were also identified using empirical Bayesian analysis (Table 1 and Supplementary Data),<sup>7</sup> revealed two interesting features. Firstly, classifying genes were found to fall within three main functional classes and secondly, these functional classes could discriminate between three molecular tumor subtypes. The first subtype showed significant increased expression of genes involved in (a) wound-healing (*ADM*, *PDGFA*, and *EFEMP2*), (b) extracellular matrix constituents and remodeling machinery (*LGALS1*, *LGALS3*, *PLAT*, *TIMP1*, and *COL5A2*), and (c) cell adhesion (*PARD3*, *DAG1*, *Kindlin1*, *ZYX*, and *ALCAM*). As all of these functions are necessary for the angiogenic properties of cells, this subtype was labeled *ANGIO* and was characteristic of the grade 4 GB samples. The next group was a mixture of histopathologic and molecular subtypes and showed increased expression of genes involved in (a) cell signaling and growth (*BMP2*, *ABII1*, *REPS2*, *ADCY2*, and *NET1*), (b) protein biosynthesis (*RPL22* and *ZMYND11*), and (c) cell cycle (*PARD3*, *ZMYND11*, and *CLASP2*). This group, labeled as *DIFFER*, characterizes the grades 2 and 3 samples, which while active in growth and neuronal differentiation have not yet acquired angiogenic properties. This group was further analyzed using a set of genes coding for ankyrin repeat proteins (*ANK3* and *ANKS1B*), solute carrier proteins (*SLCO1A2* and *SLC34A1*), a protein involved in apoptosis (*DNAJA3*), and *PEA15*, a cytostatic and antiapoptotic phosphoprotein enriched in astrocytes (22). This analysis led to the separation of the *DIFFER* group into the *INTER* subtype, which was characteristic of the grade 3 samples, and the *LOWER* subtype, which was characteristic of grade 2 samples.

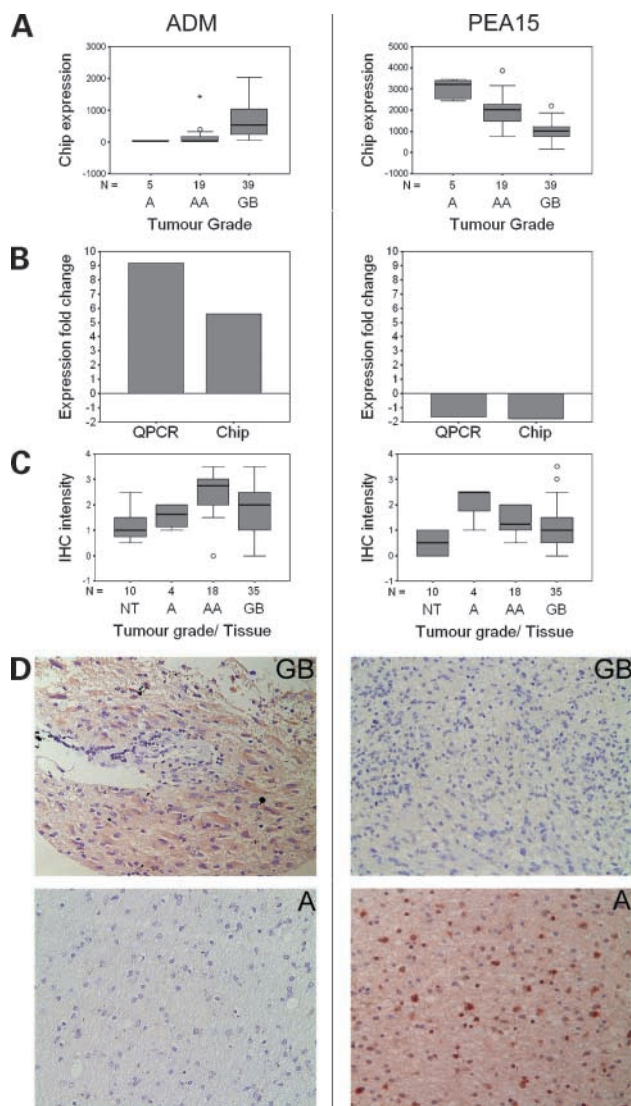
**Gene Classifiers of Particular Biological Interest.** In addition to the identification of three molecular tumor subtypes, two classifiers, *PEA15* (1q21.1, *LOWER*) and *ADM* (11p15.4, *ANGIO*), were of particular biological interest and/or novelty. These genes were also found to be differentially expressed between GB-A and/or GB-AA tumor grades by empirical Bayesian analysis. Expression changes were validated by both QPCR and immunohistochemistry (Fig. 2). A further 23 differentially expressed genes identified by Bayesian analysis were successfully validated by QPCR. The correlation ( $R^2$ ) between Affymetrix and QPCR GB/AA expression fold changes for these genes was >0.8 (data not shown).

### Grading Using Trained ANN Models

**Grading of Test Samples Using the Trained ANN Models into Tumor Subtypes Agrees with Prior Histopathologic Grading.** Grading of the test set ( $n = 26$ ) was done by passing each test sample through all models saved during the training process (for details, see Supplementary Data).<sup>7</sup> Through this way, the 59 genes/probes selected during training can be put to the test of grading a “blind” set of tumor samples. For each test sample, an initial voting was done by the *ANGIO/DIFFER* trained models. The samples that were graded as *DIFFER* received a follow-up grading by the *INTER/LOWER* trained models to discriminate between *INTER* and *LOWER* subtypes (Supplementary Table SA).<sup>7</sup> Histopathologic grading of the test samples was found to agree with the tumor subtypes observed during training. More specifically, all GB (except sample GB154) and all lower-grade astrocytic tumors (A and AA) showed increased expression of *ANGIO* and *DIFFER* genes, respectively. Furthermore, all A samples were distinguished from AA by the differential expression of *INTER/LOWER* genes. Overall, our ANN-defined tumor subtypes were in agreement with prior histopathologic grading, reporting 94.74%, 100%, and 100% accuracies for “GB,” “AA,” and “A” grading, respectively. Visualization of network outputs using available clustering algorithms (23, 24) for all train and test samples (including annotation with genomic metadata) is shown in Fig. 3A (also see Supplementary Data).<sup>7</sup>

**Grading of an Independently Published Astrocytic Tumor Gene Expression Data Set Using Cross-Chip Gene Classifiers.** To further validate the grading capacity of our gene classifiers, we used an independent, astrocytic tumor gene expression data set published by Shai et al. (9). Of the 59 probe sets selected during training from our HG-U133A GeneChips, 38 genes had >96% identity to probe sets on the U95Av2 GeneChip used by Shai et al. (9). Of these, we selected 20 genes that appeared more than once in our leave-one-out cross-validation runs, thus ensuring that only the most significant probe sets were used in the cross-chip analysis. Of these, 17 genes were differentially expressed in the GB-AA comparison (*ANGIO* and *DIFFER* genes), 1 in the GB-A comparison (*INTER/LOWER* gene), and 2 in the AA-A comparison (*INTER/LOWER* genes). We retrained the ANN models on our original training data using these





**Figure 2.** Expression of *ADM* and *PEA15* changes with astrocytic tumor progression at both transcript and protein levels. **A**, GeneChip (*chip*) expression values for *ADM* and *PEA15* across tumor grades. Samples PA68 and PA67 (grade 1 tumors; see text) have not been included in this analysis. Expression changes for both gene products are highly statistically significant (for the AA-GB and A-GB tumor grade comparisons, respectively: *ADM*,  $P = 1.1e^{-6}$  and  $1.6e^{-4}$ ; *PEA15*,  $P = 4.8e^{-5}$  and  $8.3e^{-5}$ ). **B**, validation of expression changes using QPCR. Mean expression fold changes between GB and AA tumor grades (*Expression fold change*) shown as assessed by both GeneChip and QPCR expression technology. **C**, tissue array immunohistochemistry immunoreactivity intensities for *ADM* and *PEA15* across tumor grades and nonneoplastic brain (NT, normal tissue). Five-scale grading system used: 0, no immunoreactivity; 4, intense immunoreactivity. For each tumor grade and for the collection of normal tissues, an average immunoreactivity grade was obtained from replicate tissue cores available on the tissue array. Differences in immunohistochemistry immunoreactivity (Mann-Whitney nonparametric test,  $P < 5.0e^{-2}$ ) were significant for tumor group comparisons A-GB and A-AA (*PEA15*) and A-AA (*ADM*). **D**, representative immunohistochemistry results for *ADM* and *PEA15* on GB and A tumor sections. Immunoreactivity for both gene products was seen to be present in tumor cells only. *ADM* showed cytoplasmic staining and *PEA15* showed both cytoplasmic and nuclear staining. The nuclear/cytoplasmic distribution of *PEA15* immunoreactivity was not constant for all tumor cells in a given tumor sample.

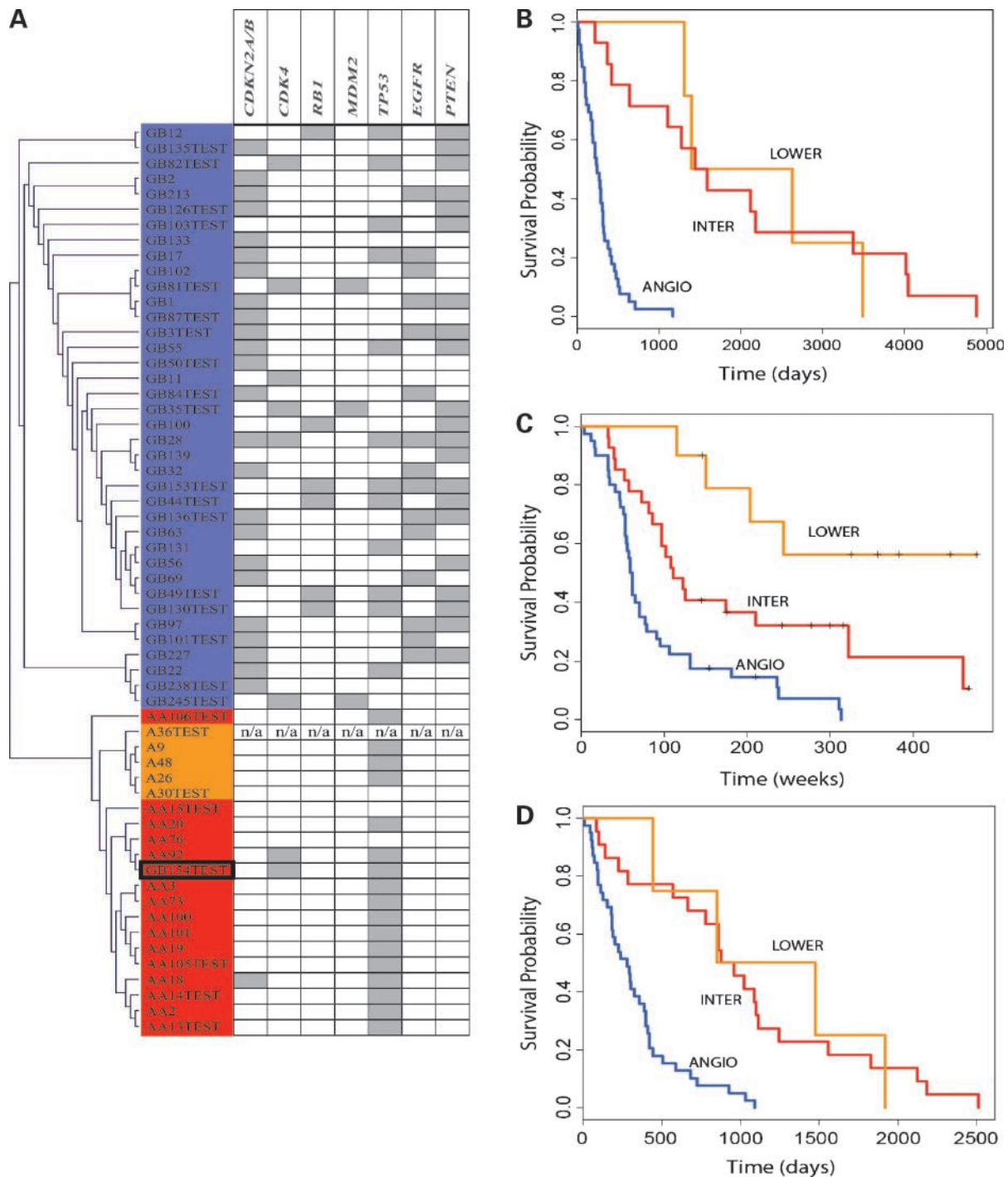
20 probe sets (for gene names, see Supplementary Data).<sup>7</sup> Due to the limited number of probe sets available for the GB-A assessment, we split the grading task into two pairwise comparisons. ANN models of “type 1” were trained to distinguish between grade 4 and lower-grade astrocytic tumors using the 17 *ANGIO/DIFFER* genes and models of “type 2” were trained to distinguish between grades 2 and 3 tumors using the three *INTER/LOWER* genes. Only samples that were graded as lower-grade *DIFFER* tumors by type 1 models required follow-up grading by type 2 models. The 23 (18 GB, 3 AA, and 2 A) samples derived from the Shai et al. data set were treated as a blind test set and were graded using our trained models. A remarkable consistency was observed between the two expression data sets using the 20 common probe sets, whereby histopathologic and ANN-based subtyping resulted in an agreement accuracy of 100% (2 of 2), 100% (3 of 3), and 88.89% (16 of 18) for the A (graded as *LOWER*), AA (graded as *INTER*), and GB (graded as *ANGIO*) tumors of the Shai et al. study, respectively (Supplementary Table SC).<sup>7</sup>

**Grading of Additional Samples Difficult to Grade Histopathologically and Evaluation of ANN Results Using Clinical, Histopathologic, and Genomic Annotation.** After verifying the grading power of our molecular signatures, we used them to identify the stage of certain samples that were particularly challenging to diagnose by histopathology. Histopathologic identification of PA (grade 1) and malignancy grading of astrocytic tumors that have been treated with irradiation and/or chemotherapy can be extremely difficult. We therefore examined the expression data from six such problem cases using the trained ANN models.

The two PA tumors (PA68 and PA67) were graded as *ANGIO* (GB-rich) and *INTER* (AA-rich), respectively, by our trained ANN models (see Discussion). These tumors were histologically typical (25) and were derived from patients with excellent survival (alive at end of follow-up; see Supplementary Table SI).<sup>7</sup> Samples AA49 and AA86 were difficult to grade as they had received irradiation and chemotherapy. Two other AA tumors, AA29 and AA93, were also difficult to grade histologically. Grading these samples using our trained ANN models did not concur with histopathologic grading and showed the grading of all 4 AA samples as *ANGIO* (GB-rich subtype; Supplementary Table SB).<sup>7</sup>

To investigate possible reasons for this discrepancy, we evaluated available annotation for all four ambiguous tumors in our data set as well as for the misgraded GB154 and the two grade 1 PA. In addition to histopathologic diagnosis, available annotation included (a) clinical data (age at operation, gender, primary or secondary tumor, and tumor location), (b) survival data, and (c) previously published genomic information for a total of nine genes (*CDKN2A*, *CDKN2B*, *p14<sup>ARF</sup>*, *CDK4*, *RB1*, *MDM2*, *EGFR*, *PTEN*, and *TP53*) known to be affected in astrocytoma (see Supplementary Table SI).<sup>7</sup>

The histology of tumors AA49 and AA86 was difficult to use for malignancy grading as previous treatment complicated the findings significantly. AA49 shows a clear GB



**Figure 3.** **A**, visualization of network results for all 33 training and 26 test tumor samples (39 GB, 15 AA, and 5 A) using the 59 genes/probe sets selected during training. Hierarchical clustering of network outputs (Euclidean distance and single linkage algorithm). The visualization represents results from the propagation of all samples (training and test) through the trained models. Color coding is according to the three molecular subtypes that best characterize the samples (*blue*, *ANGIO*; *red*, *INTER*; *orange*, *LOWER*). The only tumor where ANN subtyping does not agree with histopathology is GB154 (highlighted by a *black box*). AA106 appears to cluster separately from the rest of the AA but does not reside in the *LOWER* cluster. Tumors are annotated with genomic information for a total of seven loci available from previously published work from our laboratory (34–38), known to be involved in astrocytic tumor genesis and/or progression. *Gray boxes*, homozygous deletion (*CDKN2A/CDKN2B/p14<sup>ARF</sup>*), amplification (*CDK4*, *MDM2*, and *EGFR*), or loss of one allele with mutation of the remaining allele (*RB1*, *TP53*, and *PTEN*). *TEST*, test samples. **B**, Kaplan-Meier survival plot of our 59 astrocytic tumors as defined by our ANN grading results. *Blue line*, *ANGIO*; *red line*, *INTER*; *orange line*, *LOWER*. **C**, Kaplan-Meier survival plot of the 76 Phillips et al. samples as defined by our ANN grading results using our 59 gene classifiers. **D**, Kaplan-Meier survival plot of the 65 Freije et al. samples. *ANGIO* subtype contained 38/50 GB, *INTER* was composed of 6/15 AA and 12/50 GB, whereas the remaining 4/15 AA made up the *LOWER* subtype.

genetic profile (homozygous deletion of *CDKN2A*, *CDKN2B*, and *p14<sup>ARF</sup>*, *EGFR* amplification, and no wild-type *PTEN*), whereas AA86 shows further genetic abnormalities commonly seen in glioblastoma: lack of wild-type *CDKN2A*, *p14<sup>ARF</sup>*, or *TP53*. In the case of tumor AA29, clinical, histopathologic, and genomic evidence indicated a significant resemblance to GB (suspicion of but no frank necrosis found and no wild-type *PTEN*). Tumor AA93 had a histologic and clinical appearance of an AA but shared the same classic GB-like genetic profile seen also for AA49. The only genetic difference between the two tumors related to the retention of one wild-type copy of *PTEN*.

The four ambiguous AA tumors classified by our ANN as *ANGIO* comprised 100% (2 of 2) of the *EGFR* amplifications, 100% (2 of 2) *PTEN* mutations, and 66% (2 of 3) of the *CDKN2A/CDKN2B* nullizygosity found across all the AA samples assessed. With the exception of one *INTER* graded AA tumor with *CDKN2A/CDKN2B* nullizygosity, lesions for the cyclin inhibitor locus were totally absent in all remaining AA of our data set. All three of the AA cases, where survival data were available, died within 2 years.

No apparent reasons for the disagreement between histopathology and ANN subtyping of the GB tumor (GB154) could be found. Although GB154 had some nonclassic GB characteristics, the presence of amplification of *CDK4*, necrosis, and microvascular proliferation, the latter being major histologic criteria for glioblastoma, support the original histopathologic diagnosis. Survival in this case was also under 2 years.

#### Survival Analysis

**Survival Analysis Using the Selected Gene Classifiers Reveals a Prognostic Value for Tumor Subtypes.** To investigate the survival prognostic capabilities of our gene classifiers, we did survival analysis on our 59 samples as graded by histopathology and then as defined by our trained ANN models into the three tumor subtypes. Although there was only a small difference between ANN- and histopathology-based grading efforts (difference of one sample: GB154), the survival analysis based on the ANN grading proved to be more significant ( $P = 8.76e^{-7}$ ) than that based on purely histopathologic data ( $P = 2.088e^{-6}$ ) as defined by the log-rank test (Fig. 3B). Similar results were obtained from survival analysis of the Shai et al. (9) data set. The prognostic value of our ANN defined subtypes was equally significant ( $P = 6.0e^{-3}$ ) to that based on histopathology ( $P = 6.0e^{-3}$ ).

**Survival Analysis Substantiates Grading of Data Sets Where ANN-Defined Subtypes Do Not Concur with Prior Histopathologic Grading.** To our surprise, for two other independently published data sets, the ANN failed to recapitulate histopathologic grading. However, in both cases, survival analysis favored the ANN-based grading.

The Phillips et al. (11) data set comprised 100 MDA samples (76 for which survival information was available). In that study, the samples were divided into three "subclasses" representing the progression of astrocytic tumors. The subclasses were defined by the authors as proneural (*PN*), proliferative (*Prolif*), and mesenchymal

(*Mes*), with increasing malignancy from *PN* to *Mes*. Because those samples consisted of grades 3 and 4 tumors, we used the ANN models trained with *ANGIO/DIFFER* genes to classify the 100 MDA samples into the respective subtypes. The *ANGIO* subtype consisted of 50/76 GB and 4/24 AA, whereas the *DIFFER* subtype was comprised 22/76 GB and 12/24 AA. The *ANGIO* group consisted of 30/35 of the Phillips et al. (11) *Mes* samples, in accordance to previously published results that show that *Mes* tumors display overexpression of angiogenic markers (11). The *DIFFER* group consisted of 33/37 of the *PN* samples, also in accordance to previous reports indicating that *PN* samples display overexpression of markers of neuronal differentiation and growth. Further analysis of the *DIFFER* survival samples using our *INTER/LOWER* genes partitioned them into the *INTER* subtype, which consisted of 18/76 GB and 4/76 AA, and the *LOWER* subtype, which consisted of 4/76 GB and 8/24 AA. This approach grouped the Phillips et al. (11) samples into three very significant prognostic subclasses (Fig. 3C;  $P = 1.922e^{-7}$ ), once again outperforming the previous subtyping defined in the Phillips et al. (11) study ( $P = 1.0e^{-4}$ ). The Phillips et al. (11) *Prolif* samples, which according to their study represent the intermediate stage of the progression and are highly enriched for proliferative markers, were not so well defined by our tumor subtypes. However, 20/28 resided within the ANN-defined *ANGIO* subtype (which was rich in *Mes* samples) and 8/28 within our *INTER* subtype (rich in *PN* samples; Supplementary Table SE).<sup>7</sup> This was in accordance with previous published results that show a very similar survival median for the Phillips et al. (11) *Mes* and *Prolif* groups and a higher angiogenic index of the *Prolif* compared with the *PN* tumors (11). In addition, this concurs with the observation that the *Prolif* signature is less exclusive and the proportion of astrocytic tumors with this signature varies across samples obtained from different institutions (11).

Finally, a probe comparison between the 59 gene classifiers used in this analysis and the final 35 probes identified in Phillips et al. showed that there were no common probes between the two gene/probe sets, once again highlighting the novelty of our gene classifiers.

Similar results were obtained for another independent data set containing 65 astrocytic tumors (15 grade 3 and 50 grade 4) published by Freije et al. (4). More specifically, our subtyping significantly outperformed (Fig. 3D;  $P = 8.13e^{-8}$ ) the final survival groups obtained by Freije et al. (4) in the respective publication ( $P = 2.2e^{-4}$ ).

**Genes Predictive of Survival versus Genes Predictive of Histopathology.** To investigate this unexpected performance on the Phillips et al. and Freije et al. data sets, whereby genes identified based on histopathology acted as prognostic signatures of survival, we decided to compare genes predictive of survival (survival-correlated genes) and genes predictive of histopathology (histopathology-based genes) within our own data set as well as within the other two large data sets. We initially did clustering using the top 80 positively correlated and negatively correlated genes to



survival (Pearson's correlation of expression values versus survival times greater than 0.55 or less than -0.55) and observed three major clusters. We then recalibrated our ANN to optimize leave-one-out cross-validation runs for these clusters with survival correlated genes and resulted in an optimum set of 37 genes (see Supplementary Table SG).<sup>7</sup> We also did the same histologic-based analysis on the two independent data sets as described earlier for our own data set and selected the respective histopathology-based genes. Log-rank tests using histopathology-based or survival-correlated genes are shown in Table 2. Interestingly, we found that genes identified by signal-to-noise to be differentially expressed between histologic grades were more successful than the respective survival-correlated genes in predicting survival in all data sets tested.

**TP53 Lesions Further Separate the Grade 4 GB into Two Survival Groups.** TP53 mutations are observed in >65% of secondary GB and are considered a major hallmark that defines the separate molecular pathways, responsible for the development of the secondary GB and the primary (*de novo*) GB. To identify genes with distinct signatures for these two separate pathways, we did a leave-one-out cross-validation using only the GB separated into TP53 mutated and wild-type and identified an optimum set of 11 probes (see Supplementary Table SF).<sup>7</sup> Using these genes, the ANN separated our *ANGIO* subtype into two groups denoted as *ANGIO-PRI* and *ANGIO-SEC*. This distinction was more significant for survival prediction ( $P = 3.325e^{-2}$ ) than the respective TP53 separation ( $P = 7.082e^{-1}$ ). In the Phillips et al. data set, we found that the *ANGIO-SEC* group consisted of 16/28 *Prolif* samples and only 3/35 *Mes* samples, whereas the *ANGIO-PRI* group consisted of 12/28 *Prolif* and 32/35 *Mes* samples. This is in accordance to previous reports (26) showing that secondary GB undergo aggressive proliferation (as is the case with the *Prolif* samples) in contrast to primary GB, which show overexpression of

angiogenic genes (as is the case with the *Mes* samples). Survival analysis using our 59 gene classifiers and the 11 gene signatures described here, for all three data sets, is shown in Fig. 4.

## Discussion

In this study, we used a simple, ANN-based approach to derive specific transcriptional signatures from histopathologic subtypes of astrocytic tumors and assessed whether these molecular signatures define survival prognostic subclasses. We found that the classifier genes selected fall into three distinct functional classes, which characterize three molecular tumor subtypes, denoted *ANGIO*, *INTER*, and *LOWER*. ANN-based grading into the three tumor subtypes for our own as well as one independent data set (9) was found to accurately match prior histopathologic grading. This was not the case for two other data sets (4, 11). To investigate this discrepancy, we did an extensive comparison between survival correlated genes and histopathology-based genes. We showed that with respect to survival prediction (a) histopathology-based genes outperform the respective survival-correlated genes in each data set and (b) our histopathology-based genes outperform survival-correlated genes in *all* data sets tested. Finally, ANN analysis of TP53 mutated and wild-type samples identified a gene signature that appears to further separate the *ANGIO* subtype into two groups reflecting primary and secondary GB.

The prognostic nature of markers of angiogenesis and proliferation has been reported previously (27–30) with angiogenic markers (*VEGF*, *flt1/VEGFR1*, *kdr/VEGFR2*, and *PECAM1*) and markers of proliferation (*PCNA* and *TOP2A*) commonly used by pathologists for astrocytic tumor grading. Here, we provide a novel set of genes that characterize the *ANGIO* subtype and appear to control angiogenesis. The general trend for grade 4 GB to reside

**Table 2. Comparison of histopathology genes and genes predicted by correlation to survival**

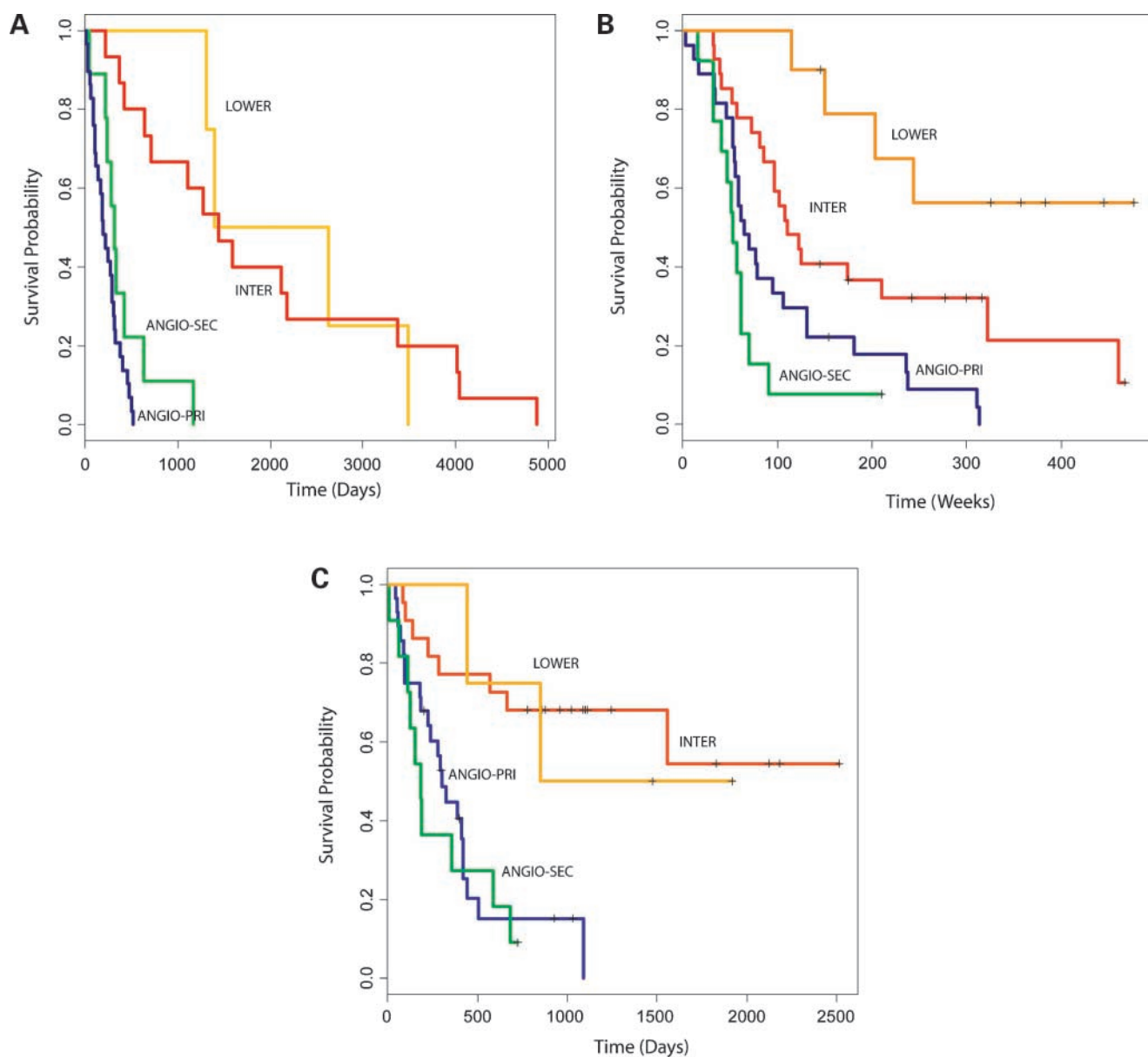
Data set	Genes used		
	Our 59 histopathology genes	Phillips et al. histopathology genes*	Freije et al. histopathology genes*
Survival ours	$P = 8.76e^{-7}$	$P = 4.258e^{-5}$	$P = 6.871e^{-6}$
Survival Phillips et al.	$P = 1.922e^{-7}$	$P = 8.808e^{-8}$	$P = 1.133e^{-4}$
Survival Freije et al.	$P = 8.13e^{-8}$	$P = 4.55e^{-8}$	$P = 2.318e^{-8}$
	Our 37 surv corr genes	Phillips et al. 35 surv corr genes <sup>†</sup>	Freije et al. 44 surv corr genes <sup>‡</sup>
Survival ours	$P = 4.377e^{-6}$	$P = 1.871e^{-4}$	$P = 4.431e^{-5}$
Survival Phillips et al.	$P = 2.081e^{-5}$	$P = 1.0e^{-4}$	$P = 3.169e^{-5}$
Survival Freije et al.	$P = 1.103e^{-5}$	$P = 1.191e^{-1}$	$P = 2.2e^{-4}$

NOTE:  $P$  values were predicted using the log-rank test after grouping the samples into survival groups.

\*These  $P$  values are the result of the grouping of the samples into two groups.

<sup>†</sup>These  $P$  values are the result of the grouping of the samples into three groups (23/35 genes were present on our GeneChip as we only used the HG-U133A GeneChip).

<sup>‡</sup>These  $P$  values are the result of the grouping of the samples into three groups (41/44 genes were present on our GeneChip as we only used the HG-U133A GeneChip).



**Figure 4.** Survival analysis of astrocytic tumors (including the 11 primary/secondary gene signature). **A**, Kaplan-Meier survival plot of our 59 astrocytic tumors as defined by our ANN grading results. *Blue line, primary ANGIO; green line, secondary ANGIO; red line, INTER; orange line, LOWER*. **B**, Kaplan-Meier survival plot of the 76 Phillips et al. samples as defined by our ANN grading results using our 59 gene classifiers + 11 gene signature. **C**, Kaplan-Meier survival plot of the 65 Freije et al. samples using our 59 gene classifiers + 11 gene signature.

within the *ANGIO* subtype is in accordance with these reports. The presence of most of the Phillips et al. (11) defined *Mes* samples within the *ANGIO* subtype further substantiates findings as these samples have been reported to overexpress angiogenic markers such as *VEGF*. The differentiating and developing nature of the lower grade AA and A is consistent with the observation that these tumors reside within the *DIFFER* group. The general trend for the Phillips et al. (11) *PN* samples to resemble the *DIFFER* samples is in accordance to reports that show that *PN* samples overexpress markers of neurogenesis and

neuronal differentiation (11). The Phillips et al. (11) *Prolif* subtype is not defined based on our tumor subtyping but was partially defined by the 11 genes used to differentiate between primary and secondary GB. The characteristic of the Phillips et al. *Prolif* samples to be less clearly defined confirms previous observations that report a less specific phenotype for these samples as well as a greater variability across samples obtained from different institutions (4, 11). Furthermore, we identified an interesting set of genes (including *PEA*) that appear to separate the *DIFFER* group (lower grade 2 A and grade 3 AA) into the *INTER* (grade 3

AA) and the *LOWER* (grade 2 A) subtypes and further define a prognostic class with the highest survival probability (*LOWER*).

Survival analysis suggests that histopathologic grading, although categorical and oversimplified, provides a general trend by which genes predictive of survival can be identified, with prognostic value greater than histopathologic grading per se. Survival prognosis can be achieved either independently or, as in the case of our data set, in conjunction with histopathology prediction. A comparison of survival-correlated and histopathology-based genes showed that the latter were more efficient in survival prognosis. This was observed for our data set as well as two other independent data sets tested. A possible explanation for this unintuitive finding relates to the methodology used to obtain survival prognostic groups. This involves the prediction of survival-correlated genes and the concurrent clustering of the tumor samples using these genes. The clusters defined are considered as prognostic groups and a unique gene signature for each cluster is obtained. This methodology is highly dependent on clustering techniques and may be less accurate than using histopathologic groups to define gene expression signatures. Other reasons include the numerous external factors that influence survival probability and do not directly relate to cancer, like the patient's age, physical and neurologic performance, etc. Genes encoding such factors will appear highly correlated with survival in small sample groups frequently used in microarray studies despite having no association with cancer per se. However, such genes may have limited predictive capacity when applied to other data sets. Expression profiles of histopathology-associated genes, on the other hand, are directly linked to cancer and are expected to be more consistent among different patients, thus having a better predictive capacity. Although there is significant variability between different studies in specimen processing, analysis, and tissue heterogeneity, which is likely to affect the identification of classifier genes, our findings show that it is possible to use expression data to identify genes with predictive capacity that extends across multiple data sets.

Two genes of special interest have been selected for further analysis in this study, that is, *PEA15* and *ADM*. Tumor-suppressing functions for *PEA15* have been suggested (31). *PEA15* suppresses DISC-mediated caspase-8 activation, limits entry to the cell cycle, and has not been associated previously with astrocytic tumor progression. Physiologic levels of *PEA15* expressed in cultured astrocytes are capable of restricting ERK to the cytosol, blocking ERK-dependent c-Fos transcription and cell proliferation (22). Candidate tumor suppressor genes, such as *PEA*, may act as major stalling points for tumor progression and perhaps the diminished expression of such genes may directly contribute to a cascade of events that lead to the progression of early-grade tumors to later more malignant phenotypes. *PEA15* was selected for further analysis to investigate its subcellular localization but also in a preliminary attempt to elucidate possible correlations

between *PEA15* expression and astrocytic tumor cell programmed cell death. *ADM* is a 52-amino acid peptide suggested to be capable of affecting tumor growth by both direct tumor cell-related mitogenic effects and indirect vasculature-related angiogenic mechanisms (32). *ADM* expression in astrocytic tumors has been shown previously, whereas its increased expression with tumor grade progression was recently suggested by Tso et al. (26). *ADM* was selected to validate previous suggestions relating the peptide to regulation of angiogenesis and because very few publications commented on its exact subcellular or tissue localization.

This work presents a large, new expression profiling data set of astrocytic tumors and employs a novel ANN-based grading of these tumors into molecular subtypes. We show that it is possible to derive transcriptome signatures from the tripartite histopathologic grading used to train the ANN-model. Moreover, these signatures attain a more significant survival prognosis when compared with histopathologic grading as well as tumor subtyping reports from other studies. We hope that the identification of the novel set of genes underlying this subtyping will enable tumor diagnosis to progress toward a more quantitative realm, where tumors are viewed within a malignancy spectrum that includes samples from all stages of tumor progression. We also believe that the interpretation of grading and classification efforts based on gene expression data must be done using thorough tumor annotation on as many levels as possible. It is the integration of such work with clinical, genotypic, and histopathologic annotation that can maximize the value of gene expression data, increase our understanding of tumor pathology, and further develop current diagnostic and therapeutic approaches.

#### Acknowledgments

We thank François Renault-Mihara, INSERM, Chaire de Neuropharmacologie, Paris, France, for his most generous *PEA15* antibody gift.

#### References

- Kleihues P, Cavenee WK. Pathology and genetics of tumours of the nervous system. Lyon: IARC Press; 2000.
- Ichimura K, Ohgaki H, Kleihues P, Collins VP. Molecular pathogenesis of astrocytic tumours. *J Neurooncol* 2004;70:137–60.
- Mischel P, Cloughesy T, Nelson S. DNA-microarray analysis of brain cancer: molecular classification for therapy. *Nat Rev Neurosci* 2004;10:782–92.
- Freije WA, Castro-Vargas FE, Fang Z, et al. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* 2004;64:6503–10.
- Nutt CL, Mani DR, Betensky RA, et al. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* 2003;63:1602–7.
- Liang Y, Diehn M, Watson N, et al. Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci U S A* 2005;102:5814–9.
- Kim S, Dougherty ER, Shmulevich I, et al. Identification of combination gene sets for glioma classification. *Mol Cancer Ther* 2002;1:1229–36.
- van den Boom J, Wolter M, Kuick R, et al. Characterization of gene expression profiles associated with glioma progression using oligonucleotide-based microarray analysis and real-time reverse transcription-polymerase chain reaction. *Am J Pathol* 2003;163:1033–43.
- Shai R, Shi T, Kremen TJ, et al. Gene expression profiling identifies molecular subtypes of gliomas. *Oncogene* 2003;22:4918–23.

10. Rickman DS, Bobek MP, Misk DE, et al. Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer Res* 2001;61:6885–91.
11. Phillips HS, Kharbanda S, Chen R, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 2006;9:157–73.
12. Ekstrand AJ, James CD, Cavenee WK, Seliger B, Pettersson RF, Collins VP. Genes for epidermal growth factor receptor, transforming growth factor  $\alpha$ , and epidermal growth factor and their expression in human gliomas *in vivo*. *Cancer Res* 1991;51:2164–72.
13. Ihaka R, Gentleman R. R: a language for data analysis and graphics. *J Comput Graph Stat* 1996;5:299–314.
14. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
15. Barrett T, Suzek TO, Troup DB, et al. NCBI GEO: mining millions of expression profiles-database and tools. *Nucleic Acids Res* 2005;33: Database Issue: D562–6.
16. Hosack D, Dennis G, Sherman B, Lane H, Lempicki R. Identifying biological themes within lists of genes with EASE. *Genome Biol* 2003;4:R70.
17. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 2001;25: 25:402–8.
18. Oehler MK, Fischer DC, Orlowska-Volk M, et al. Tissue and plasma expression of the angiogenic peptide adrenomedullin in breast cancer. *Br J Cancer* 2003;89:1927–33.
19. Sharif A, Renault F, Beuvon F, et al. The expression of PEA-15 (phosphoprotein enriched in astrocytes of 15 kDa) defines subpopulations of astrocytes and neurons throughout the adult mouse brain. *Neuroscience* 2004;126:263–75.
20. Kaplan EMP. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457–81.
21. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–7.
22. Renault F, Formstecher E, Callebaut I, Junier M-P, Chneiweiss H. The multifunctional protein PEA-15 is involved in the control of apoptosis and cell cycle in astrocytes. *Biochem Pharmacol* 2003;66:1581–8.
23. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–25.
24. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863–8.
25. Burger PC, Scheithauer BW. Atlas of tumor pathology. Washington (DC): Armed Forces Institute of Pathology; 1994.
26. Tso CL, Freije WA, Day A, et al. Distinct transcription profiles of primary and secondary glioblastoma subgroups. *Cancer Res* 2006;66: 159–67.
27. Ho DM, Hsu CY, Ting LT, Chiang H. MIB-1 and DNA topoisomerase IIa could be helpful for predicting long-term survival of patients with glioblastoma. *Am J Clin Pathol* 2003;119:715–22.
28. Hsu SC, Volpert OV, Steck PA, et al. Inhibition of angiogenesis in human glioblastomas by chromosome 10 induction of thrombospondin-1. *Cancer Res* 1996;56:5684–91.
29. Osada H, Tokunaga T, Nishi M, et al. Overexpression of the neuropilin 1 (NRP1) gene correlated with poor prognosis in human glioma. *Anticancer Res* 2004;24:547–52.
30. Godard S, Getz G, Delorenzi M, et al. Classification of human astrocytic gliomas on the basis of gene expression: a correlated group of genes with angiogenic activity emerges as a strong predictor of subtypes. *Cancer Res* 2003;63:6613–25.
31. Gaumont-Leclerc MF, Mukhopadhyay UK, Goumard S, Ferbeyre G. PEA-15 is inhibited by adenovirus E1A and plays a role in ERK nuclear export and Ras-induced senescence. *J Biol Chem* 2004;279:46802–9.
32. Benes L, Kappus C, McGregor GP, Bertalanffy H, Mennel HD, Hagner S. The immunohistochemical expression of calcitonin receptor-like receptor (CRLR) in human gliomas. *J Clin Pathol* 2004;57:172–6.
33. Saeed A, Sharov V, White J, et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003;34: 374–8.
34. Ichimura K, Bolin MB, Goike HM, Schmidt EE, Moshref A, Collins VP. Deregulation of the p14ARF/MDM2/p53 pathway is a prerequisite for human astrocytic gliomas with G<sub>1</sub>-S transition control gene abnormalities. *Cancer Res* 2000;60:417–24.
35. Reifenberger G, Reifenberger J, Ichimura K, Meltzer PS, Collins VP. Amplification of multiple genes from chromosomal region 12q13-14 in human malignant gliomas: preliminary mapping of the amplicons shows preferential involvement of CDK4, SAS, and MDM2. *Cancer Res* 1994;54: 4299–303.
36. Ichimura K, Schmidt EE, Goike HM, Collins VP. Human glioblastomas with no alterations of the CDKN2A (p16INK4A, MTS1) and CDK4 genes have frequent mutations of the retinoblastoma gene. *Oncogene* 1996;13: 1065–72.
37. Liu L, Ichimura K, Pettersson EH, Goike HM, Collins VP. The complexity of the 7p12 amplicon in human astrocytic gliomas: detailed mapping of 246 tumors. *J Neuropathol Exp Neurol* 2000;59: 1087–93.
38. Schmidt EE, Ichimura K, Goike HM, Moshref A, Liu L, Collins VP. Mutational profile of the PTEN gene in primary human astrocytic tumors and cultivated xenografts. *J Neuropathol Exp Neurol* 1999;58: 1170–83.