

Marked Genomic Differences Characterize Primary and Secondary Glioblastoma Subtypes and Identify Two Distinct Molecular and Clinical Secondary Glioblastoma Entities

Elizabeth A. Maher,^{1,4} Cameron Brennan,¹⁰ Patrick Y. Wen,^{1,4} Laura Durso,¹ Keith L. Ligon,⁵ Aaron Richardson,¹ Deepak Khattry,² Bin Feng,² Raktim Sinha,² David N. Louis,¹¹ John Quackenbush,^{3,7} Peter McL. Black,⁸ Lynda Chin,^{2,4,6} and Ronald A. DePinho^{2,4,9}

¹Center for Neuro-Oncology, ²Center for Applied Cancer Science, Belfer Institute for Innovative Cancer Science, ³Departments of Biostatistics and Computational Biology and Cancer Biology, Dana-Farber Cancer Institute; ⁴Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School; Departments of ⁵Pathology and ⁶Dermatology, Brigham and Women's Hospital and Harvard Medical School; ⁷Department of Biostatistics, Harvard School of Public Health; ⁸Department of Neurosurgery, Brigham and Women's Hospital; ⁹Department of Medicine and Genetics, Harvard Medical School, Boston, Massachusetts; ¹⁰Neurosurgery Service, Memorial Sloan-Kettering Cancer Center, Weill-Cornell Medical College, New York, New York; and ¹¹Department of Pathology, Molecular Pathology Unit, CNY7, Massachusetts General Hospital and Harvard Medical School, Charlestown, Massachusetts

Abstract

Glioblastoma is classified into two subtypes on the basis of clinical history: "primary glioblastoma" arising *de novo* without detectable antecedent disease and "secondary glioblastoma" evolving from a low-grade astrocytoma. Despite their distinctive clinical courses, they arrive at an indistinguishable clinical and pathologic end point highlighted by widespread invasion and resistance to therapy and, as such, are managed clinically as if they are one disease entity. Because the life history of a cancer cell is often reflected in the pattern of genomic alterations, we sought to determine whether primary and secondary glioblastomas evolve through similar or different molecular pathogenetic routes. Clinically annotated primary and secondary glioblastoma samples were subjected to high-resolution copy number analysis using oligonucleotide-based array comparative genomic hybridization. Unsupervised classification using genomic nonnegative matrix factorization methods identified three distinct genomic subclasses. Whereas one corresponded to clinically defined primary glioblastomas, the remaining two stratified secondary glioblastoma into two genetically distinct cohorts. Thus, this global genomic analysis showed wide-scale differences between primary and secondary glioblastomas that were previously unappreciated, and has shown for the first time that secondary glioblastoma is heterogeneous in its molecular pathogenesis. Consistent with these findings, analysis of regional recurrent copy number alterations revealed many more events unique to these subclasses than shared. The pathobiological significance of these shared and subtype-specific copy number alterations is reinforced by their frequent occurrence, resident genes with clear links to cancer,

recurrence in diverse cancer types, and apparent association with clinical outcome. We conclude that glioblastoma is composed of at least three distinct molecular subtypes, including novel subgroups of secondary glioblastoma, which may benefit from different therapeutic strategies. (Cancer Res 2006; 66(23): 11502-13)

Introduction

The recalcitrant nature of glioblastoma, despite the application of novel therapies targeting "signature glioma genes" in clinical trials, continues to motivate efforts to define the genetics and biology of this aggressive brain tumor. The lethal nature of glioblastoma relates to its propensity to invade early and widely throughout the normal brain parenchyma and to its intense apoptosis resistance rendering it incurable with existing chemo/radiotherapies. Glioblastoma can develop along two distinct clinical pathways: primary glioblastoma, which presents with no pathologic precursor lesion and has a clinical prodrome of <3 months, and secondary glioblastoma, which develops over a 5- to 10-year period from a low-grade astrocytoma (1). Molecular analysis of the glioblastoma subtypes has identified epidermal growth factor receptor (*EGFR*) amplification as exclusive to primary glioblastoma and *TP53* loss as the genetic hallmark of low-grade astrocytoma and secondary glioblastoma (1), and transcriptional differences seem to reflect biological differences between these clinical subtypes (2). At the same time, the strikingly similar phenotype of the glioblastoma subtypes is reflected in common genetic lesions such as loss of phosphatase and tensin homologue (*PTEN*) and cyclin-dependent kinase (CDK) inhibitor 2A (*CDKN2A*), as well as amplification/overexpression of *CDK4* and *MDM2* (1). Emerging data, however, suggest that these genetic events may represent only a fraction of mutations involved in glioblastoma pathogenesis.

Conventional and array-based comparative genomic hybridization (aCGH) and microarray expression profiling have catalogued the common chromosomal aberrations, each of which contains genes with copy number-driven expression changes validated across large glioblastoma sample sets (3–5). Application of higher resolution aCGH platforms has identified additional loci targeted for recurrent copy number alterations (CNA), underscoring the genomic complexity of this aggressive cancer. Attempting to ascribe biological significance to these various loci, several studies

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Current address for E.A. Maher: Annette G. Strauss Center for Neuro-Oncology, Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390-9186.

R.A. DePinho is an American Cancer Society Professor and an Ellison Medical Foundation Scholar.

Requests for reprints: Ronald A. DePinho, Department of Medical Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115. Phone: 617-632-6085; Fax: 617-632-6069; E-mail: ron_depinho@dfci.harvard.edu.

©2006 American Association for Cancer Research.
doi:10.1158/0008-5472.CAN-06-2072

have correlated changes in copy number or expression with prognosis. Whereas gain of 7p has been linked to poor prognosis (6), *EGFR* amplification and activated *EGFR* mutant (VIII) expression have not been shown to be consistent independent predictors of survival in glioblastoma (7, 8). Long-term survival (>3 years) has been correlated with 19q loss, whereas 6q loss, 10q loss, and 19q gains have been correlated with shorter survival (9)—the clinically relevant gene targets of most of these regions have yet to be identified. In the case of 10q, whereas loss of *PTEN* has not been shown to predict survival time in glioblastoma (8), coexpression of *PTEN* and activated *EGFR* was correlated with a good response to *EGFR* inhibition in a recurrent glioblastoma clinical trial (10).

Within glioblastoma expression datasets, gene clusters (5) and genes (4) correlate with survival and seem to have relevance to glioma biology and heterogeneous cellular morphology (11). Integrating aCGH and expression data into survival analysis has identified molecular subsets in glioblastoma, thus enabling a more focused search for genes involved in glioblastoma pathogenesis and clinical outcome (4). Such studies underscore the value of genomic analysis of clinically annotated tumor samples and, at the same time, emphasize the limited scope of knowledge surrounding the identity and function of the presumed cancer genes targeted in these loci.

High-resolution oligo-based aCGH platforms and advances in computational methods provide an opportunity to obtain a detailed view of the genomic changes in cancer and to correlate such CNA patterns with the tumor subtypes and clinical outcome. In this study, the application of this approach to clinically annotated, formalin-fixed paraffin-embedded primary and secondary glioblastoma samples readily documented all known glioblastoma CNAs and, in the vast majority of cases, narrowed significantly the common regions targeted for amplification or deletion. Unsupervised classification, including genomic nonnegative matrix factorization (gNMF) methods, defined primary and secondary glioblastomas as different genetic diseases and identified two novel genetic subgroups in secondary glioblastoma. These genomic profiles point to prime therapeutic candidates but emphasize that effective treatment strategies will likely require tailored and multiple targeted therapies for each of the three glioblastoma molecular subtypes.

Materials and Methods

Identification of primary tumors. Tumors were identified through the Neuro-Oncology Program at Dana-Farber Cancer Institute/Brigham and Women's Hospital and included review of case records to classify tumors by histopathologic diagnosis and clinical history under an Institutional Review Board–approved protocol. Primary glioblastomas had no prior history of glioma, had a clinical prodrome of ≤ 3 months, and were diagnosed as glioblastoma at the first operation. Secondary glioblastoma patients had a histologic diagnosis of low-grade astrocytoma that preceded the diagnosis of glioblastoma by at least 1 year. Slides from each case were reviewed by a neuropathologist (K.L.L.) to confirm the diagnosis of glioblastoma and low-grade astrocytoma, when relevant. To study as homogeneous subgroups as possible, tumors with histologic evidence of a prominent oligodendroglial component were excluded given the known genotypic differences between oligodendroglial and astrocytic tumors (12). Similarly, the “giant cell” variant of glioblastoma was excluded because it has genetics that have been described as overlapping primary and secondary glioblastomas (1). For each case, gender, age, location of tumor, treatment course, clinical course, including time to tumor transformation/progression, and survival time were collected.

DNA extraction. The details of the extraction method are described in Supplementary Methods. Briefly, one representative block per case containing >90% cellularity and minimal necrosis was selected. Twenty-five milligrams of tissue were obtained in 40- μ m histologic sections. High-quality DNA (>10 kb fragments) was extracted by modifications of Qiagen (Valencia, CA) tissue extraction protocol. Formalin-fixed paraffin-embedded sections were deparaffinized in a series of three steps: treatment with mixed xylenes, removal of xylene with 100% ethanol, and rehydration in PBS. Samples were lysed in buffer and proteinase K and inverted at 55°C for 24 to 72 hours and prepared using the Qiagen protocol. DNA quality was examined by gel electrophoresis and samples with at least 50% of total DNA running above the 10-kb marker (New England Biolabs, Ipswich, MA) were of sufficient quality to proceed with aCGH (Supplementary Fig. S1A).

aCGH profiling on oligonucleotide microarrays. Genomic DNA was fragmented and random prime labeled as previously described (13) and hybridized to oligonucleotide arrays containing 22,500 elements designed for expression profiling (Human 1A V2, Agilent Technologies, Santa Clara, CA). Using National Center for Biotechnology Information Build 35, 16,097 unique map positions were defined with a median interval between mapped elements of 54.8 kb. Fluorescence ratios of scanned images were calculated as the average of two paired arrays (dye swap), and the raw profiles were processed to identify statistically significant transitions in copy number using Circular Binary Segmentation (13, 14).

Definition of thresholds for analysis. A “segmented” data set was generated by determining uniform copy number segment boundaries and then replacing raw log₂ ratio for each probe by the mean log₂ ratio of the segment containing the probe. The distribution of segment mean log₂ ratios for the 37 samples is shown in Supplementary Fig. S1B. The tall central peak represents the majority of the genome, which is without CNA in the sample set. Based on this distribution, thresholds representing minimal CNA were chosen at ± 0.15 , equal to ± 6 SDs of the middle 75% of the data (blue lines). Thresholds representing more significant CNAs were chosen at ± 0.4 , representing the 97% and 4% quantiles (green lines). High-priority minimal common region (MCR) were chosen by requiring at least one sample to show an extreme CNA event, defined by thresholds +0.8 and -0.65 , >99% and <1% quantiles, respectively (red lines).

Genomic NMF analysis and Fisher's exact test of aCGH profiles. Genomic NMF was applied to the current data sets as previously described (15). Briefly, the segmented data set was first dimension reduced by eliminating redundant probe locations and then transformed to nonnegative values. The resultant data set is a nonnegative matrix that is subject to gNMF using a custom software package (16) and run in MATLAB (The MathWorks, Inc., Natick, MA). For each factor level 2 through 6, gNMF is repeated 1,000 times to build a consensus matrix, and this is used to assign samples to clusters based on the most common consensus. The rank $K = 3$ clustering was further significance tested by permuting sample labels for secondary glioblastoma samples independently for each chromosome. One hundred permutations were subject to rank 3 NMF over 1,000 iterations and the consensus matrix assessed by cophenetic correlation.

Fisher's exact test was used to identify significantly different regional gains or losses between primary and secondary glioblastomas. For each aCGH probe, each sample was classified as having copy number normal, gained or lost based on log₂ ratio thresholds of ± 0.15 (minimal CNA threshold; see above). Two-by-two contingency tables tested gained versus normal and lost versus normal between primary and secondary glioblastomas. Fisher's exact test P values were corrected for multiple testing (q -value FDR 10%, “ q value” package for R).¹²

Quantitative PCR verification. PCR primers were designed to amplify products of 100 to 150 bp within target and control sequences as described (13).

Automated MCR definition. Loci of amplification and deletion were evaluated across the subgroups in an effort to define MCRs targeted by

¹² <http://cran.r-project.org>.

overlapping events in two or more samples. An algorithmic approach is described in Supplementary Methods and has previously been described (13, 15).

Comparison with published expression data (KLL/DNL expression). Expression data from ref. 17 were used to generate a comparison data set of gene expression changes in glioblastoma versus normal brain ($n = 8$ samples). Statistical significance of overexpression or underexpression in glioblastoma relative to normal brain was set at $P < 0.05$.

Results

Recurrent Genomic Changes Characterize the Glioblastoma Clinical Subtypes

Array-CGH methods, employing a high-density gene-specific oligo-based platform (see Materials and Methods), were used to detect CNAs in the genomes of clinically annotated and pathologically verified primary and secondary glioblastoma cases using DNA extracted from formalin-fixed paraffin-embedded tissue (see Materials and Methods). An approximately equal number of primary ($n = 20$) and secondary ($n = 17$) glioblastomas, all of pure astrocytic histology and secondary glioblastomas arising from histologically verified low-grade astrocytomas, were included in this study to have statistical power to detect differences between the two clinical subgroups. The cases are representative of patients treated in academic neuro-oncology centers in which the median age at diagnosis is significantly less for secondary glioblastoma [40 years (range, 19-54 years), versus 51 years (range, 33-65 years) for primary glioblastoma; $P < 0.0001$] but the overall response to radiation and cytotoxic chemotherapy is similar. Median overall

survival time was similar between the groups, 17 months (range, 10-47 months) for primary glioblastoma and 14 months (range, 2-60 months) for secondary glioblastoma (primary versus secondary, $P = 0.95$; Supplementary Fig. S2). All except one patient with secondary glioblastoma were treated with external beam radiotherapy at the time of low-grade astrocytoma diagnosis and 13 of 17 (76%) were treated with adjuvant nitrosurea-based chemotherapy or temozolomide. Time to development of secondary glioblastoma was not correlated with any clinical feature or treatment modality (data not shown).

This established high-resolution aCGH platform (18) and modified segmentation analysis (ref. 13, see Materials and Methods) readily identified large regional changes of low amplitude, hereafter referred to as "gain" or "loss" (i.e., segment mean \log_2 ratios of >0.15 and ≤ 0.15), as well as focal higher-amplitude CNAs representing more significant amplification or deletion events (i.e., average size spans 2.78 Mb; range, 0.06-20.23 Mb) and \log_2 ratio amplitude ranges from +0.4 to +4.6 and -0.4 to -3.1 (see Materials and Methods). Skyline recurrence plots of primary and secondary glioblastoma profiles mirror well the frequencies of known glioblastoma loci targeted for copy number alteration (refs. 1, 19; Fig. 1) and readily show regions of gain and loss that are distinct between these classic subtypes. Ch7 gain and ch10 loss were identified in $>90\%$ of the primary glioblastomas in this sample set. Whereas it has been shown that primary glioblastoma can be divided based on the presence or absence of these chromosomal changes (20), the high resolution of the oligo-based platform as well as the sample set design (both primary and secondary

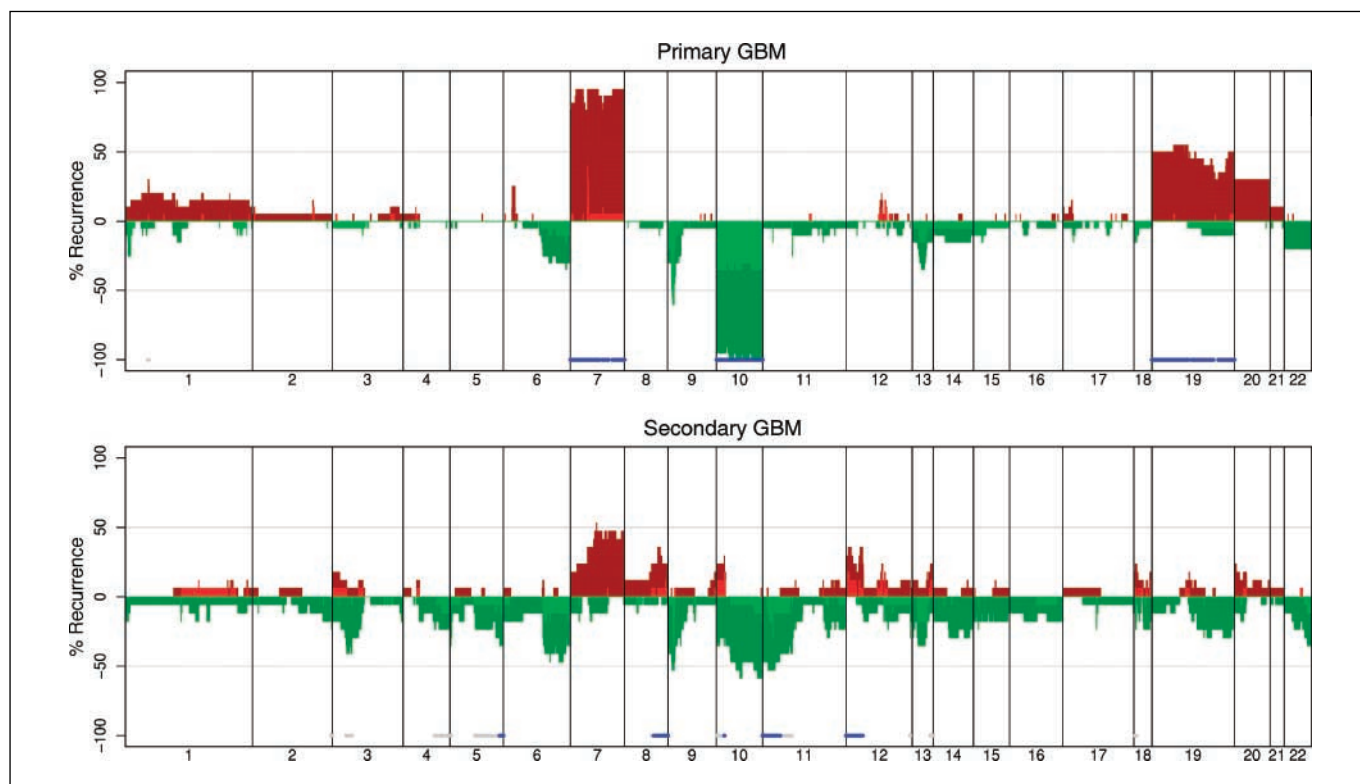


Figure 1. Recurrence plots for primary and secondary glioblastomas. Percent recurrence of CNAs across the samples in segmented data (y axis) is plotted for each aCGH probe evenly aligned along the x axis in chromosomal order. Dark red or green bars, percent of samples with approximate gain or loss of chromosomal material, as defined in Materials and Methods; bright red or green bars, number of samples showing approximate amplification or deletion (Materials and Methods). Blue bars, regions that are differentially altered between primary and secondary glioblastomas (Fisher's exact test, corrected for multiple testing by q -value, FDR 15%). Gray bars, regions marginally significant by P value ($P < 0.05$, FDR $>15\%$).

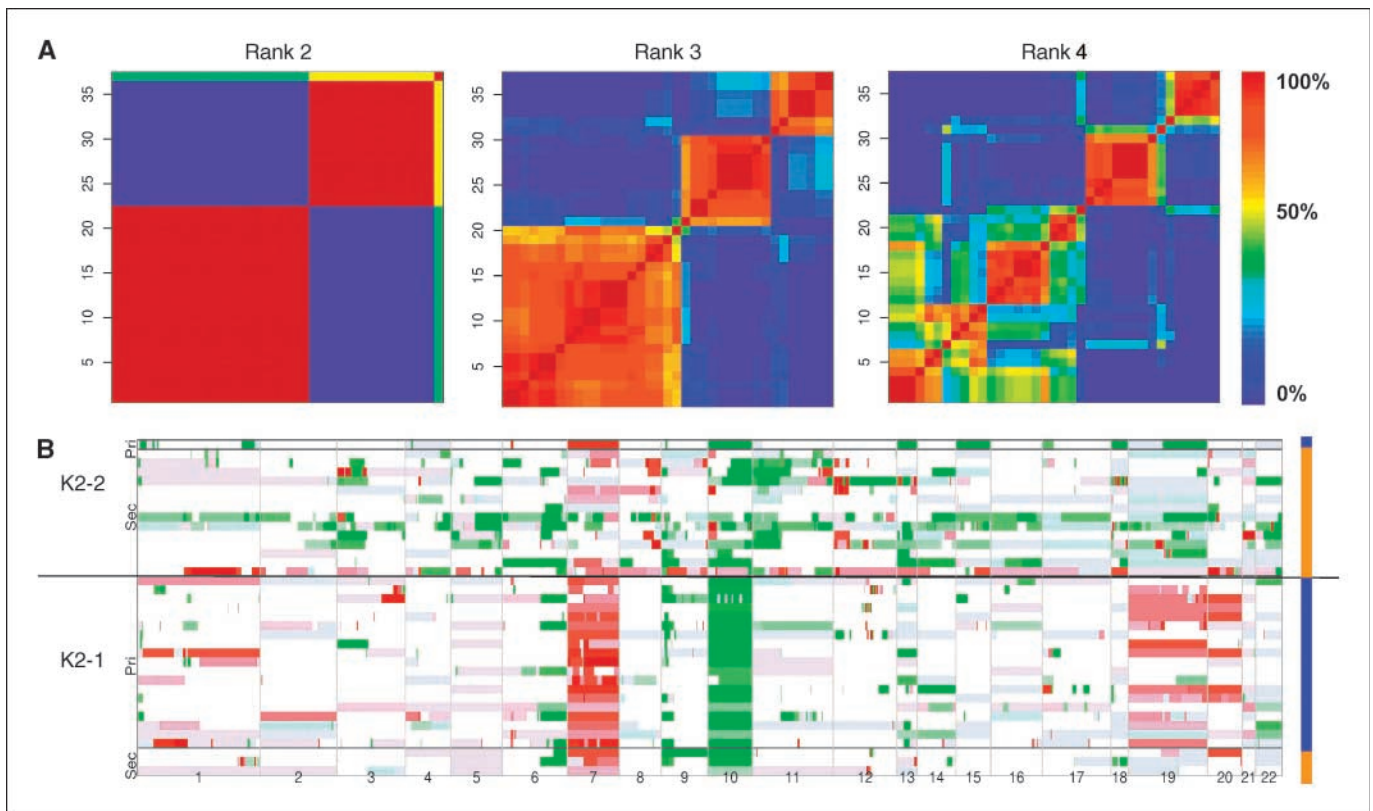


Figure 2. Primary and secondary glioblastomas clustered by gNMF. *A*, gNMF of aCGH profiles finds stable clustering for ranks 2 and 3 but not for ranks 4 or higher, suggesting up to three natural subgroups (*color plots*). Consensus matrices show how often samples are assigned to the same clusters during 1,000 repetitions of NMF, computed at $K = 2$ to 8 for 37 primary and secondary glioblastoma aCGH profiles. Each pixel represents how often a particular pair of samples cluster together, colored from 0% (*deep blue*, samples are never in the same cluster) to 100% (*dark red*, samples are always in the same cluster). Ranks 2 and 3 show largely stable assignments into 2 and 3 blocks, respectively, whereas rank 4 is disrupted (*top*). *B*, unsupervised clustering of aCGH profiles by gNMF, Rank $K = 2$, largely recapitulates clinically defined primary and secondary glioblastoma subgroups. Heat map plot shows genomic gain/amplification (*red*) and loss/deletion (*green*) for each sample. Three secondary and one primary samples are clustered to contrary groups.

glioblastomas) in this study allowed for identification of many distinct regions associated with the primary and secondary glioblastoma subtypes. Gain of ch19, in addition to ch7 gain and ch10 loss, was significantly overrepresented in primary compared with secondary glioblastomas ($P < 0.01$, FDR 10%) whereas secondary glioblastoma showed unique gains in subregions of 8q, 10p, and 12p ($P < 0.01$, FDR 10%). Loss of 9p and 10q was common for primary and secondary glioblastomas, reflecting the classic loss of *p16* and *PTEN* in the two subtypes. However, loss in other chromosomal regions was overrepresented among secondary glioblastomas and reached statistical significance for 5q, 11p ($P < 0.01$, FDR 10%), 3p, and 4q ($P < 0.017$, FDR 15%), all well delineated by the array.

Unsupervised Classification of Glioblastoma Genome Recapitulates Clinical Subgroups

Historically, glioblastoma subclassification has been based entirely on clinical history. Here, we sought to determine if chromosomal aberration patterns define genomic subclasses that corroborate or expand the classic clinical subclasses. We used gNMF, an unsupervised clustering method designed to extract distinctive genomic features from aCGH profiles and proven useful for identifying novel clinical subtypes in multiple myeloma (15, 16). After transforming the segmented aCGH data to nonnegative values, gNMF consensus matrices were generated for ranks 2 to 8,

representing attempts to cluster into 2 to 8 separate groups. Ranks $K = 2$ and 3 showed relatively stable cluster assignments, whereas groupings of rank ≥ 4 were unstable (Fig. 2A). The stability of these ranks was confirmed by a strong cophenetic correlation value of ~ 0.99 for ranks 2 and 3 but an abrupt decrease to 0.94 with $K = 4$ (Supplementary Fig. S3).

The rank 2 classification (K2) essentially encapsulated the primary and secondary glioblastoma clinical classes (Fig. 2B), 19 of 20 primary glioblastomas (identified by blue bar to right) being assigned to one cluster (K2-1) and 14 of 17 secondary glioblastomas (orange bar) assigned to the other (K2-2). The concordance between unsupervised classification and clinical class was better than that achieved by unsupervised hierarchical clustering, which was driven predominantly by gain of ch7 and loss of ch10 in the primary glioblastomas, consistent with what has been described for the genetic subgrouping of primary glioblastoma (20), as well as loss of ch19 in the secondary glioblastomas (Supplementary Fig. S4). Alteration of chromosome 19 shows an additional insight made by applying gNMF to our data set. As reflected in the skyline recurrence plot (Fig. 1), it has been established that chromosome 19 can be gained or lost in glioblastoma. Using gNMF, we show that gain of ch19 is frequent and exclusive to primary glioblastoma and, unexpectedly, tracks closely with gain of ch20 (Fig. 2B). In contrast, loss of ch19 is a feature of the K2-1 genetic subgroup, almost exclusive to secondary

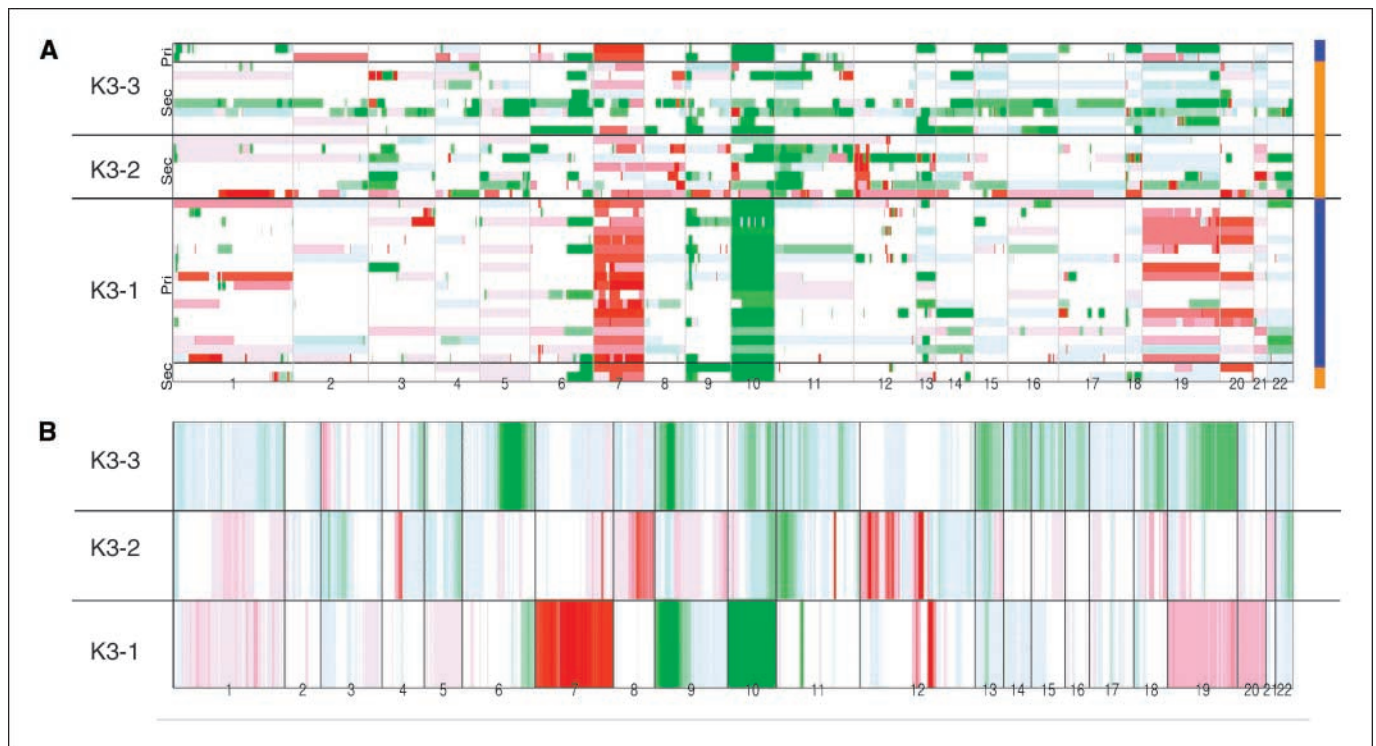


Figure 3. NMF using rank $K = 3$. *A*, heat map plot (see Fig. 2) for rank 3 clustering shows preservation of primary/secondary division while also identifying two subgroups of secondary glioblastoma. Secondary glioblastoma clusters are significant by permutation testing ($P < 0.01$; see Results). *B*, the individual components from NMF, $K = 3$, converted to gain/loss representation. Components represent the CNA pattern that defines cluster assignment, with gain/loss shown as red/green, respectively.

glioblastoma, with gain of ch20 rarely seen in this subgroup. The strong correlation between genomic profile and clinical class strongly supports the view that the pathogenesis of these classic clinical subgroups are driven by a large number of distinct genetic events.

The rank $K = 3$ classification largely preserved the K2-1 cluster of primary glioblastomas (K3-1, with 18 of 20 primary glioblastoma; *blue bar*, Fig. 3*A*). Viewing the individual gNMF components (Fig. 3*B*) indicates that assignment to K3-1 is driven by gains

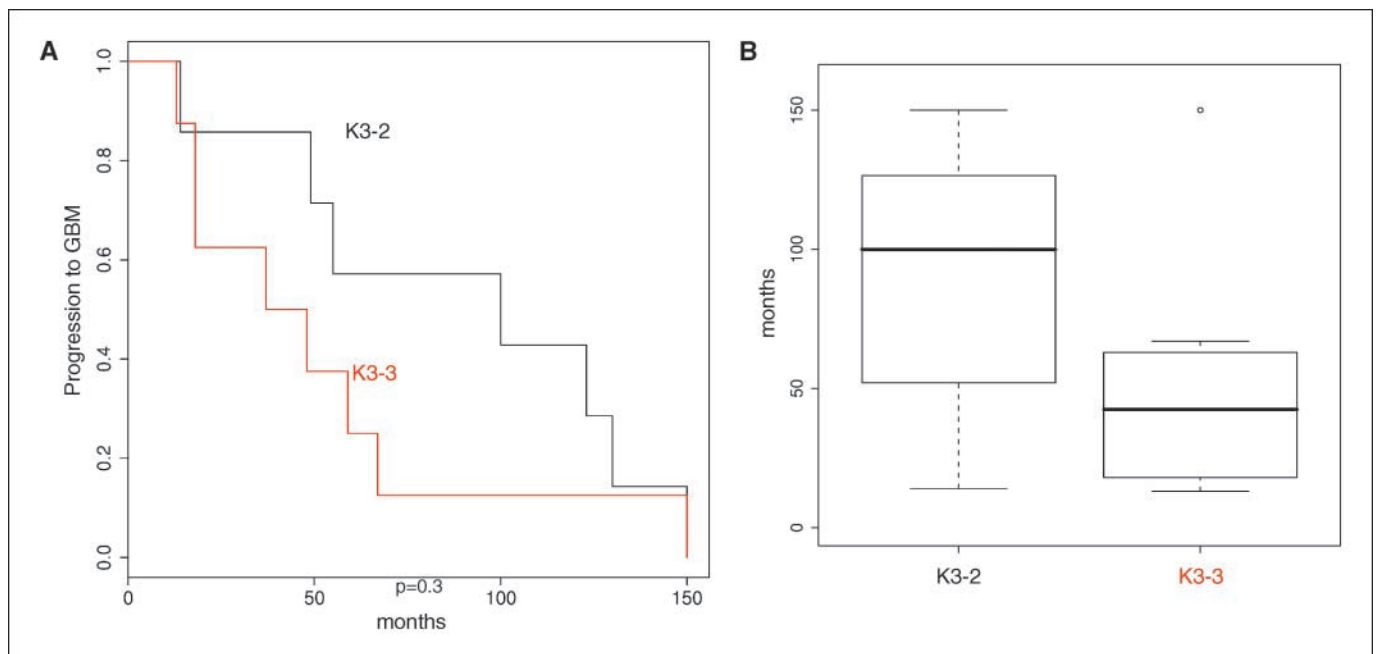


Figure 4. Time to glioblastoma progression (*TTP*) from low-grade astrocytoma in the two secondary glioblastoma gNMF subgroups. *A*, Kaplan-Meier curve. *B*, box plot, with K3-3 outlier shown excluded from range bars.

of chromosome 7, 19, and 20 and a small region of chromosome 12, as well as loss of chromosome 10 and smaller regions of 9 and 11, features typical of primary glioblastoma. Notably, the K3 classification reveals two distinct subclasses of secondary glioblastoma: K3-3 is dominated by regions of loss, in particular on chromosomes 6, 9, 10, 13, 18, and 19, whereas K3-2 has prominent regional gains on chromosomes 4, 8, and 12 and focal gains on ch7 and ch11.

The cluster stability for rank $K = 3$ gNMF strongly supports the presence of these two molecularly distinct subclasses of secondary glioblastoma in our data set, each with signature alterations correlated across the genome. To address the statistical significance of the K3-2 and K3-3 gNMF subclassification, random permutation ($n = 100$) of the sample labels for each chromosome was done (see Materials and Methods). This procedure maintains the distribution of CNAs among secondary glioblastoma samples overall while breaking correlation between chromosomes. The cophenetic correlations for rank 3 calculated for each of 100 iterations showed a mean of 0.918 (range, 0.839-0.965). Thus, the high correlation for rank 3 in the initial dataset (0.99) reflects significantly nonrandom correlation of CNA events defining the K3 secondary glioblastoma subclasses.

A further support for significance of these gNMF subclasses derives from our assessment of whether these two genetic subclasses of secondary glioblastoma were associated with any clinical variables. Whereas these divisions did not track with any of the baseline patient characteristics or treatment modality of antecedent low-grade astrocytoma, they seemed to correlate with time to progression of secondary glioblastoma from LGA disease by Kaplan-Meier analysis (Fig. 4). Whereas statistically significant differences were not achieved with this limited sample size (log-rank test, $P = 0.31$), it is worth noting a trend toward shorter time to tumor progression in the K3-3 group (42 months) versus the K3-2 (100 months). In summary, the stratification of secondary glioblastoma into two genomic subclasses was statistically significant in our dataset and the classes may have further biological significance in predicting time to tumor progression. This shows that despite their shared histopathology, primary and secondary glioblastomas are molecularly distinct and that secondary glioblastoma consists of two molecular entities apparently driven by radically different pathogenetic mechanisms.

Presence of Many Recurrent and Novel Amplifications and Deletions in Glioblastoma

The gNMF clustering of genomic profiles is dominated by recurrent CNA events, which often span long regions such as chromosomes or chromosomal arms. Whereas such broad gains and losses provide useful classifying features, the more focal events revealed by the high-resolution platform in this study may hold valuable information about tumor subclass and the target genes driving these diseases. For example, whereas complete or partial gain of ch7 is not uncommon in secondary glioblastomas (10 of 17), the discrete amplification of the region containing EGFR is seen frequently and exclusively in primary glioblastomas.

We first sought to generate a complete catalog of discrete CNAs in glioblastoma, without classification by clinical history. Narrow events were catalogued using a previously described approach to defining MCRs of overlapping CNAs among sample sets (13, 15). Analysis of the 37 independent primary and secondary glioblastoma samples defined 222 MCRs of gain/amplification or loss/deletion based on data smoothed by segmentation (Supplementary Table S1). To identify MCRs with potentially greater pathogenetic

significance, we further applied a criterion that at least one CNA of an MCR exceeded log 2 thresholds of +0.8 for amplifications and -0.65 for deletions. A total of 97 "high-priority" MCRs fulfilled this criterion (Supplementary Table S2), 41 amplifications with a median size of 0.92 Mb (range, 0.1-14.51 Mb) containing a median number of 8 known genes (range, 2-60) and 56 deletions with a median size of 1.13 Mb (range, 0.07-22.38 Mb) containing a median of 10 known genes (range, 2-233). High confidence is ascribed to these MCRs as evidenced by (i) consistent verification via real-time quantitative PCR in 34 of 36 (94%) randomly assayed high-priority MCRs (data not shown), and (ii) complete overlap with known glioblastoma loci including focal deletions encompassing *CDKN2A*, *P TEN*, and *RB* and focal amplifications targeting *EGFR*, platelet-derived growth factor receptor α (*PDGFR\alpha*), *CDK4*, *MDM2*, *MET*, and *c-MYC* (Supplementary Table S1). Notably, 17 of 97 MCRs contain hotspots for proviral integration in diverse cancer models including a murine glioma model (21) and 12 MCRs contain microRNAs (Supplementary Table S2), several of which are overexpressed in glioblastoma (22).

Highly Focal Amplifications and Deletions with Cancer-Relevant Genes

The resolution of the platform employed enabled precise definition of MCR boundaries, revealing 42 MCRs that were present in ≥ 4 (10%) tumors and spanning ≤ 1.0 Mb with a median of five known genes. Examination of these resident genes revealed known oncogenes, tumor suppressor genes, or their highly related homologues, most of which have not been ascribed to gliomagenesis but many that contain genes with strong cancer relevance. Some notable genes in amplicons include *EPST15* (1p32), *WNT7A* (3p26.3-3p24.3), and *RASSF3*, a RAS association family member (12q14.2), and in loss/deletion MCRs *IRF2BP2*, an IFN regulatory factor 2 repressor (1q42.2) and *MKK4/JNKK1* (6q26). Even more striking are the profiles of 15 MCRs spanning < 0.5 Mb (median, 4 genes) and present in > 4 (10%) tumor samples (Supplementary Table S2, bold and italics). Among these, some notable candidates not previously ascribed to glioblastoma include extracellular signal-regulated kinase (ERK)-8 (8q23-24 amplicon), a novel member of the RAS-mitogen-activated protein kinase family that is activated by a SRC-independent mechanism, and *TMPO/LAP2* (12q22 deletion), a protein that anchors Rb in the nucleoskeleton and regulates Rb function. Recognizing that the sample size is limited, it is intriguing that deletions of *TMPO/LAP2* and RB are mutually exclusive.

Comparison of Primary and Secondary Glioblastoma Genomic Profiles

The differences observed in the skyline recurrence plots of primary and secondary glioblastomas (Fig. 1) and the gNMF classification (Fig. 2) prompted a comparison of MCRs of primary and secondary glioblastomas to identify focal common and subtype-specific genomic events that may drive pathogenesis. Primary glioblastomas contain 101 MCRs (38 amplifications, 63 deletions) with a median size of 1.28 Mb (range, 0.07-24.9 Mb; Supplementary Table S3), and secondary glioblastomas contain 135 MCRs (52 amplifications, 83 deletions) with a median size of 2.94 Mb (range, 0.06-21.03 Mb; Supplementary Table S4). Only 59 MCRs were shared, of which 24 seem to be novel loci and 35 reside within known glioblastoma loci, although it is worth noting that many of these previously reported loci are very large and our MCRs have narrowed the regions of interest considerably (Table 1). Strikingly, most MCRs proved to be subtype specific (see below).

Table 1. MCRs common to primary and secondary glioblastoma

Cytogenetic band	Locus (Mb)	Minimal common region				MCR recurrence			
		Position (Mb)	Size (Mb)	Peak value	No. transcripts	Gain loss (%)	Amp/del (%)	Known novel	Verified by quantitative PCR
Gain and amplification									
1p31.1-1p22.3	76.09-86.92	76.09-86.92	10.82	0.46	47	13.5	5.4	Novel	
1q32.1-1q32.1	198.16-198.57	198.16-198.57	0.41	1.00	4	16.2	5.4	Known	NAV1
1q32.1-1q32.2	200.47-203.98	200.47-203.98	3.51	1.86	59	13.5	2.7	Known	GAC1
1q44-1 q44	240.54-241.85	240.54-241.85	1.31	1.90	11	13.5	2.7	Novel	
3p26.3-3p24.3	0.42-21.45	10.95-11.57	0.62	1.18	5	10.8	5.4	Novel	
3p26.3-3p24.3	0.42-21.45	12.91-13.83	0.92	1.18	5	10.8	5.4	Novel	HDAC11 and WNT7A
4p12-4p12	47.43-47.90	47.43-47.90	0.47	0.96	4	5.4	2.7	Novel	
4q11-4q13.1	52.80-62.59	53.45-56.07	2.63	2.43	11	8.1	5.4	Known	PDGFR α
5q31.2-5q31.2	138.98-139.89	138.98-139.89	0.91	0.49	11	10.8	2.7	Novel	
7p11.2-7p11.2	54.41-56.43	54.41-55.26	0.85	4.39	4	75.9	37.8	Known	EGFR (PGBM only)
7q22.1-7q32.1	101.70-126.60	115.79-121.29	5.50	2.70	18	77.8	18.9	Known	MET, CAV1 (SGBM only)
9p21.2-9p13.2	26.83-37.62	32.44-32.53	0.09	0.93	3	10.8	5.4	Known	
9q34.3-9q34.3	135.62-136.57	135.62-136.57	0.95	0.42	18	13.5	2.7	Novel	MRPL41; CACNA1B
12q13.2-12q13.3	54.51-55.27	54.88-55.27	0.39	0.69	17	10.8	8.1	Known	RAP1B
12q13.3-12q14.1	55.31-57.55	56.21-56.28	0.07	3.64	3	32.4	24.3	Known	CDK4
12q14.3-12q15	65.02-66.88	65.02-66.88	1.85	2.61	6	8.1	5.4	Novel	DYRK2
12q15-12q15	66.92-69.80	67.00-67.95	0.95	4.28	10	24.3	21.6	Known	MDM2
15q23-15q24.1	68.12-72.25	71.38-71.79	0.41	0.67	6	16.2	8.1	Novel	
15q26.3-15q26.3	99.66-100.20	99.66-100.20	0.54	0.60	5	10.8	2.7	Novel	
19p13.12-19q13.11	15.77-40.20	19.75-24.13	4.38	0.77	33	56.8	8.1	Novel	GPATC1, SLC7A10 (in locus)
Loss and deletion									
1p36.33-1p36.22	1.55-9.35	7.83-9.12	1.28	-0.68	14	24.3	13.5	Known	
1p31.1-1p22.2	77.73-89.03	78.12-87.52	9.40	-0.48	41	13.5	5.4	Novel	
1q32.2-1q42.12	203.98-223.12	214.89-220.65	5.76	-0.63	31	18.9	10.8	Known	
1q43-1q44	236.39-245.42	244.52-244.85	0.32	-1.19	14	10.8	10.8	Novel	OR2M7 and OR2T4
2q37.1-2q37.1	233.25-234.03	233.25-234.03	0.78	-0.60	10	13.5	5.4	Novel	TNRC15 and INPP5D
4p13-4p12	44.01-48.72	47.77-48.72	0.95	-0.55	10	8.1	5.4	Novel	
5p15.33-5p15.31	0.23-9.68	0.91-3.65	2.74	-0.51	19	24.3	8.1	Known	
6p25.3-6p22.3	0.29-17.71	0.29-0.35	0.06	-1.33	2	21.6	8.1	Novel	DUSP22
6q16 2-6q21	99.42-107.22	105.28-105.83	0.55	-0.65	5	40.7	24.1	Known	
8p22-8p12	13.30-30.04	16.01-17.13	1.12	-0.78	6	13.5	8.1	Known	
9p24.3-9p21.1	0.20-30.54	21.06-21.99	0.93	-2.30	22	55.6	46.3	Known	p16
10q22.1-10q24.31	74.43-102.75	84.73-89.59	4.86	-1.38	28	85.2	46.3	Known	PTEN
10q22.1-10q24.31	74.43-102.75	90.68-91.05	0.37	-2.13	5	81.5	44.4	Known	
10q22.1-10q24.31	74.43-102.75	92.49-102.73	10.24	-0.64	104	81.5	44.4	Known	
10q24.32-10q25.2	104.38-112.03	104.38-112.03	7.65	-0.64	36	79.6	40.7	Known	MXI1 and others
10q25.2-10q26.3	112.04-135.24	120.43-121.38	0.95	-1.27	9	83.3	48.1	Known	GPR10 and GRK5
10q25.2-10q26.3	112.04-135.24	123.71-129.78	6.08	-1.27	48	85.2	48.1	Known	GPR10 and GRK5
10q25.2-10q26.3	112.04-135.24	133.87-135.02	1.16	-1.27	17	85.2	48.1	Known	
11p15.4-11p15.4	5.71-5.82	5.73-5.79	0.07	-3.11	6	29.6	16.7	Known	
11q11-11q11	55.07-55.29	55.09-55.19	0.09	-2.43	3	29.6	13.0	Known	OR4C11 and OR4S2
11q11-11q12.3	55.32-63.09	57.34-61.91	4.57	-0.47	114	18.9	5.4	Known	LOC390183
11q14.1-11q23.1	82.21-112.62	85.82-89.56	3.73	-0.58	20	24.3	8.1	Known	GPR83
11q14.1-11q23.1	82.21-112.62	93.55-94.44	0.89	-0.93	10	14.8	10.8	Known	GPR84
11q14.1-11q23.1	82.21-112.62	95.20-102.06	6.87	-0.62	19	27.0	8.1	Known	
12p13.2-12p13.1	11.31-14.54	12.37-12.83	0.46	-0.62	7	10.8	5.4	Novel	LOH1CR12
12q21.1-12q21.31	73.73-80.15	73.73-75.65	1.92	-0.65	10	16.7	8.1	Novel	
12q22-12q24.23	94.19-118.99	97.43-97.49	0.07	-0.65	2	13.5	8.1	Known	
12q22-12q24.23	94.19-118.99	105.88-112.29	6.40	-0.51	81	16.2	8.1	Novel	
12q24.32-12q24.32	127.71-131.09	130.94-131.07	0.11	-0.51	3	10.8	5.4	Novel	
13q12.11-13q13.3	19.14-35.64	19.66-20.64	0.99	-0.60	11	32.4	18.9	Known	
14q11.2-14q11.2	19.40-19.55	19.40-19.51	0.11	-0.68	3	33.3	37.8	Known	
15q13.3-15q15.3	30.71-41.87	40.43-41.02	0.59	-0.68	8	22.2	16.2	Novel	ZFP106, TTBK2, FLJ23375
15q25.2-15q26.3	80.20-97.49	80.20-88.17	7.97	-0.57	73	24.3	8.1	Novel	

(Continued on the following page)

Table 1. MCRs common to primary and secondary glioblastoma (Cont'd)

Cytogenetic band	Locus (Mb)	Minimal common region				MCR recurrence			
		Position (Mb)	Size (Mb)	Peak value	No. transcripts	Gain loss (%)	Amp/del (%)	Known novel	Verified by quantitative PCR
17q21.2-17q21.2	35.77-36.97	36.59-36.66	0.07	-2.45	4	18.5	13.5	Known	
18p11.32-18p11.22	0.20-10.66	6.95-10.66	3.71	-0.70	16	27.0	18.9	Known	
19q13.2-19q13.43	47.16-63.76	47.78-48.55	0.77	-0.97	19	24.3	10.8	Known	
19q13.2-19q13.43	47.16-63.76	53.06-54.83	1.76	-0.59	84	24.3	10.8	Known	AURKC (and others)

Abbreviations: PGBM, primary glioblastoma; SGBM, secondary glioblastoma.

Shared MCRs. Among the 35 shared known loci (10 amplifications and 25 deletions) were deletion of *RB*, *CDKN2A*, and *PTEN*, and amplification of *CDK4*, *MDM2*, *c-MYC*, and *GAC1*. Of interest, the 4q11-13 amplification containing *PDGFR α* was present in both subtypes. Twenty-two shared novel loci (11 amplifications, 11 deletions) were quite focal—0.98 Mb median size, median 10 resident genes. Although not previously described in glioblastoma, there is a high degree of confidence that these MCRs are cancer relevant because a significant number of the MCR resident genes in the amplicons exhibit increased expression in glioblastoma relative to normal brain ($P < 0.05$; see Materials and Methods and Supplementary Table S2) and/or these loci and resident genes have been implicated in the pathogenesis of other cancer types. For example, the ErbB receptor ligand *NRG2* (neuregulin-2), which resides in a 5q MCR amplicon, is overexpressed in glioblastoma compared with normal brain ($P < 0.05$, KLL/DNL dataset) and has been correlated with node positivity and worse outcome in breast cancer (23). The same MCR contains *DTR* (heparin binding EGF-like growth factor), which is up-regulated in inflammatory breast cancer and associated with poor prognosis (24). Another 12q13-15 amplicon with four MCR genes contains the dual specificity kinase *DYRK2*, which is amplified and/or overexpressed in gastrointestinal stromal tumor, esophageal cancer, and non-small-cell lung cancer (25). An intriguing MCR spanning ~4 Mb at 19p13.12, which is gained/amplified in 40% of glioblastoma cases, is particularly notable for the presence of many genes almost exclusively encoding zinc finger proteins.

Similarly, recurrent deletions point to a number of prime candidates that are targeted in both subtypes. For example, ~10% of glioblastomas sustain deletion of 2q37.1, which contains *INPP5D/SHIP1*, an inositol polyphosphate-5-phosphatase family member that has been implicated in leukemia and gastrointestinal stromal tumor (26). Focal loss on 12q24 contains three genes, *SFRS8* (a splicing factor for fibronectin, among others), *MMP17* (MT4-MMP), a membrane-type matrix metalloproteinase that is down-regulated in prostate cancer (27), and *ULK1*, a novel serine-threonine kinase (28).

Subtype-specific MCRs. Whereas subtype-specific differences are well established for a handful of loci, 60% of the MCRs were specific to either primary or secondary glioblastoma (Supplementary Tables S3 and S4, respectively). It is worth noting that this degree of difference is comparable to that reported for pancreas versus lung cancer (13), further substantiating the distinct pathogenesis of the classic glioblastoma subtypes. For primary

glioblastoma-specific MCRs, there are 23 gains/amplifications with a median size of 0.87 Mb and median number of 13 genes. Of these 23 loci, 6 represent loci that have been significantly narrowed and 11 loci have not been described in glioblastoma, yet their cancer relevance is reinforced by validated connections to other cancer types such 2p25 in ganglioneuroblastoma (29), 6p25 in uveal melanoma (30), and 17q21 in medulloblastoma (ref. 31; Table 2). For the loss/deletion MCRs, there are 35 loci with a median size of 1.86 Mb and median of 16 genes. For most loci, the MCR represents a significantly narrowed focal region(s) within large regions or whole chromosomal arms of loss of heterozygosity (LOH) previously described in glioblastoma. There is strong cancer relevance within these narrowed loci, with many being associated with multiple diverse cancers; some examples include loss of 3q12 in small-cell lung cancer (32), leukemia (33), and nasopharyngeal carcinoma (34); loss of 4q34 in gastric cancer (35), hepatocellular cancer (36), and primitive neuroectodermal tumor (37); and loss of 15q21 in bladder cancer (38) and rhabdomyosarcoma (39). Thus, it is likely that these candidate tumor suppressor genes in primary glioblastoma bear relevance to other cancer types.

For secondary glioblastoma-specific MCRs, there are 34 gains/amplification (median size 3.25 Mb, median of 27 genes). Consistent with previous reports of *MET* amplification occurring independently of *EGFR* amplification (40), we observed *MET* amplification only in secondary glioblastoma. Of the remaining 33 loci, none have previously been described in secondary glioblastoma and only a few have been found in glioblastoma datasets not otherwise classified. However, several loci have been reported at least by band location in other cancer types [e.g., 11q13 in head and neck squamous cell cancer (41) and 20q12-13 in poor prognosis esophageal cancers (42)], supporting the cancer relevance of these subtype specific loci. With regard to the loss/deletion MCRs, there are 54 loci with median size of 3.02 Mb and median of 27 genes. Most loci have either not been described for secondary glioblastoma or have been significantly narrowed (e.g., 19q13.12-q13.2). Commonly, large regions of LOH that are reported for glioblastoma have been narrowed in this analysis to small regions that are distinct from primary glioblastoma. Again, the potential relevance of these loci to glioblastoma pathogenesis is reflected by their association with other cancer subtypes [e.g., loss of 1q23 in medulloblastoma (43), 5q11 in head and neck squamous cell cancer (44), and 8p12 in hepatocellular carcinoma (45)].

Among the large number of subtype-specific MCRs, 18 of 58 primary glioblastoma-specific MCRs fulfill the high-priority

Table 2. High-priority glioblastoma subtype-unique MCRs

Cytogenetic band	Locus (Mb)	Minimal common region			
		Position (Mb)	Size (Mb)	Peak value	Transcripts
Primary glioblastoma					
1p32.3-1p32.3	51.13-52.21	51.13-52.21	1.08	1.03	12
2q33.1-2q33.2	201.36-203.68	201.6-201.92	0.32	0.95	5
6p21.2-6p21.2	37.71-39.11	37.71-39.11	1.39	0.85	5
7p14.1-7p14.1	38.04-38.2	38.04-38.2	0.15	1.16	23
7p12.3-7p12.1	49.72-50.86	50.28-50.86	0.57	4.27	4
7q22.1-7q22.1	100.03-100.14	100.03-100.14	0.10	0.87	6
9q33.2-9q33.3	122.66-122.97	122.66-122.97	0.31	1.06	8
19q13.33-19q13.33	53.8-53.96	53.8-53.96	0.15	0.84	13
1p36.11-1p35.3	27.62-28.65	27.62-28.65	1.02	-0.62	16
1q42.2-1q42.3	230.11-231.05	230.11-231.05	0.94	-1.02	5
4q33-4q34.1	171.01-175.63	171.01-175.63	4.61	-0.67	15
6p21.32-6p21.32	32.47-32.72	32.47-32.72	0.25	-0.81	7
6p12.1-6p12.1	53.89-55.73	53.89-55.73	1.83	-0.63	9
6q21-6q21	109.41-109.88	109.41-109.88	0.46	-0.63	8
6q25.3-6q25.3	158.06-160.18	158.06-160.08	2.01	-0.70	17
9p13.3-9p13.2	35.04-38.54	35.72-35.9	0.17	-0.72	8
9q21.11-9q21.2	69.23-77.27	69.23-77.27	8.02	-0.67	29
10q11.23-10q22.1	51.26-71.86	69.66-70.64	0.98	-0.80	15
11q13.5-11q14.1	77.26-78.05	77.26-78.05	0.78	-0.80	9
17q21.31-17q21.33	40.54-46.3	44.04-44.3	0.25	-0.92	7
21p11.2-21q22.11	9.93-31.11	14.4-15.26	0.85	-0.69	7
Secondary glioblastoma					
1q42.13-1q44	224.91-245.43	231.04-234.33	3.27	1.62	22
3p24.1-3p21.32	29.91-44.43	29.91-44.43	14.51	0.94	96
8q23.1-8q24.3	109.86-146.26	144.87-146.26	1.38	1.76	55
10p15.3-10p11.21	0.28-35.97	30.64-35.97	5.33	0.84	18
11q13.5-11q14.1	77.26-78.05	77.26-78.05	0.78	3.40	9
12p13.33-12p13.2	0.17-12.15	3.81-6.68	2.86	2.45	38
12p12.3-12q12	15.02-37.98	21.59-29.56	7.96	2.56	37
12q13.13-12q13.13	52-52.16	52-52.16	0.15	4.55	5
12q14.2-12q14.2	62.87-63.4	62.87-63.4	0.52	1.92	6
12q21.1-12q21.2	73.73-74.71	73.73-74.71	0.98	1.13	6
12q24.33-12q24.33	130.94-131.08	130.94-131.08	0.12	1.26	3
18q11.2-18q11.2	17.33-20.9	17.33-20.9	3.56	0.81	21
18q12.1-18q12.3	26.82-39.11	26.82-39.11	12.27	0.82	44
19q12-19q13.2	34.79-43.49	35.73-38.96	3.22	0.86	23
1p13.3-1p13.3	107.82-108.96	107.82-108.96	1.13	-0.99	8
2q36.1-2q36.1	223.26-224.57	223.26-224.57	1.30	-0.76	7
3p24.3-3p24.3	19.55-20	19.55-20	0.44	-1.64	4
5q23.1-5q33.3	118.2-156.28	134.21-140.59	6.37	-0.67	104
6q14.1-6q15	79.7-88.06	79.7-88.06	8.34	-0.76	30
6q23.3-6q23.3	136.59-138.25	136.59-138.25	1.65	-0.76	11
9p21.2-9p21.2	27.51-27.94	27.51-27.94	0.42	-0.86	4
10p15.3-10p13	0-28-12.91	0.28-3.82	3.53	-0.62	15
11q23.2-11q25	113.77-133.78	127.83-133.78	5.93	-0.69	34
12q14.1-12q14.1	57.55-58.47	57.55-58.47	0.91	-0.76	2
12q14.2-12q15	63.17-67.45	63.39-67.45	4.05	-0.65	27
12q24.33-12q24.33	129.17-129.39	129.17-129.39	0.21	-0.64	2
12q24.33-12q24.33	131.09-132.35	131.09-132.35	1.25	-0.64	19
14q13.1-14q13.2	33.97-34.95	33.97-34.95	0.97	-0.62	13
14q32.12-14q32.31	90.79-101.47	90.79-101.47	10.67	-0.67	100
14q32.32-14q32.33	102.03-106.34	102.03-106.34	4.29	-0.67	233
16q23.1-16q23.2	73.98-79.15	73.98-79.15	5.16	-0.71	18
18q12.3-18q21.31	30.57-54.22	40.89-50.12	9.21	-0.73	43
20p13-20p11.21	0.02-25.61	7.9-18.72	10.81	-0.85	61
20p13-20p11.21	0.02-25.61	19.96-24.6	4.63	-0.73	44

Table 2. High-priority glioblastoma subtype-unique MCRs (Cont'd)

MCR recurrence				Viral integration sites	MicroRNA	Candidate genes
Gain/loss		Amp/del				
T	%	T	%			
3	9.1	1	3.0			<i>EPS15</i>
4	12.1	3	9.1			
3	9.1	1	3.0			
29	87.9	4	12.1			
27	81.8	6	18.2			
27	81.8	2	6.1			
2	6.1	1	3.0			
12	36.4	1	3.0			
3	9.1	1	3.0			
4	12.1	2	6.1			
2	6.1	2	6.1			
3	9.1	2	6.1			
5	15.2	2	6.1			
12	36.4	7	21.2			
15	45.5	10	30.3	NKI-34		
10	30.3	3	9.1			
3	9.1	3	9.1	Anxa1	hsa-mir-204	
31	93.9	19	57.6			
5	15.2	2	6.1			
4	12.1	3	9.1		hsa-mir-196a-1	
5	15.2	2	6.1			
2	9.5	2	9.5	D8Erttd233e		
2	9.5	1	4.8	CKLF7	hsa-mir-128b, hsa-mir-26a-1	<i>MLH1, CTNNB1</i> <i>RECQL4</i>
7	33.3	5	23.8			
5	23.8	3	14.3	MAP3K8		
2	9.5	1	4.8			
6	28.6	6	28.6	CCND2		<i>ZNF384</i>
6	28.6	5	23.8	KRAS2, LRMP, STK381, BHLHb3		<i>KRAS2</i>
2	9.5	1	4.8			
3	14.3	1	4.8			
2	9.5	1	4.8			
2	9.5	1	4.8			
2	9.5	2	9.5	Evi3	hsa-mir-133a-1, hsa-mir-1-2	
2	9.5	2	9.5	MAPRE2	hsa-mir-187	
4	19.0	2	9.5			<i>CEBPA</i>
3	14.3	1	4.8			
3	14.3	2	9.5			
5	23.8	3	14.3			
5	23.8	5	23.8	KLH-L3		
8	38.1	8	38.1	A930014C21Rik		
11	52.4	10	47.6			
8	38.1	4	19.0			
11	52.4	6	28.6			<i>COPEB</i>
6	28.6	4	19.0	Fli1 Ets1		<i>FLII</i>
2	9.5	2	9.5			
6	28.6	3	14.3	RAP1B		<i>HMGA2</i>
3	14.3	2	9.5			
4	19.0	2	9.5			
4	19.0	2	9.5	NFKBIA		
7	33.3	5	23.8	4831426/19Rik	hsa-mir-370, hsa-mir-299	<i>TRIP11, GOLGA5, TCL1A, BCL11B</i>
6	28.6	4	19.0	IGH4, AKT1	hsa-mir-203	
4	19.0	3	14.3			<i>MAF</i>
5	23.8	4	19.0	4933427L07Rik		
3	14.3	3	14.3	SNX5		
3	14.3	3	14.3			

criteria and several are very focal—8 amplifications and 10 deletions with median sizes of 0.31 and 0.89 Mb. In the secondary glioblastoma-specific MCRs, 34 of 88 are high-priority MCRs (Table 2) and 5 amplifications and 5 deletions are highly focal (<1.0 Mb) with a median number of 5 genes, including an amplicon at 11q13.5-11q14.1 that contains *GAB2*, a docking protein involved in EGF-induced ERK activation (46).

Discussion

In this study, gNMF-based unsupervised classification of glioblastoma genomic profiles distinguished the known clinical subgroups and identified previously unappreciated wide-scale genomic differences between primary and secondary glioblastomas. We have further identified novel subclasses of secondary glioblastoma grounded in clinical history. In addition to many classic loci, we identified more than 20 novel loci common to all glioblastoma subtypes as well as numerous novel loci specific to primary and secondary glioblastomas. These subcategories of MCRs, linked to clinical and potentially biologically important subclasses of glioblastoma, provide a molecularly grounded framework for the study of glioblastoma pathogenesis and for the more rapid prioritization of genes for validation and ultimately drug discovery. On a more technical level, our ability to use archival tissue to generate DNA of quality sufficient for analysis by high-resolution aCGH has enabled the use of clinically annotated tumor specimens that, without frozen tissue, have largely been inaccessible to gene-specific whole genome profiling for basic investigation and discovery.

Using the clinical information to generate subgroups of tumors for analysis proved useful in suggesting new pathogenetic pathways and may be useful in guiding the application of targeted therapies. The unsupervised clustering of glioblastoma into three genetic subgroups suggests distinct routes of pathogenesis and tumor biology and raises questions about whether a common phenotype results from functional redundancy within varied genetic pathways or from the existence of critical common overlap genes that ultimately determine phenotype. The answers to these questions are of prime relevance to the development of effective therapies for glioblastoma because establishing what specific genes, gene patterns, and/or genetic pathways are responsible for the uniformity of biological and clinical outcome will ultimately identify prime targets for glioblastoma therapy. Along these lines, it is worth recalling the recent clinical trials of EGFR inhibitors for recurrent glioblastoma. Whereas fewer than 10% of patients derived benefit from the EGFR inhibitors (47), there is clear evidence that clinical response requires normal PTEN activity and constitutive EGFR activation (10). On examination of our glioblastoma genomic profiles, one would predict infrequent response to EGFR inhibitors. Specifically, whereas primary glioblastoma commonly harbors amplified *EGFR* (with approximately half of these expressing an activated mutant allele), the majority of our cases also sustained loss of chromosome 10 encompassing *PTEN*. Along similar lines, examination of secondary glioblastoma profiles shows neither *EGFR* amplification nor activating mutations, strongly suggesting that these patients would fail to respond to EGFR inhibition regardless of PTEN status. It stands to reason that, because treatment planning and clinical trial design have been based on a small number of signature mutations, the more complete genomic profiles of this study could provide drug response biomarkers for EGFR and other targeted therapies.

A recent study has proposed that the genomic differences between primary and secondary glioblastomas may be reflected at a transcriptional level as well (2). This study has reported only 21 genes distinctly overexpressed in secondary and 58 genes in primary glioblastoma, when each glioblastoma group was compared with grade 2/3 astrocytoma. An additional 15 genes were overexpressed in both primary and secondary glioblastomas. Whereas this study did not conduct direct comparisons of primary and secondary glioblastoma expression patterns, they provided the opportunity for comparison of these expression changes with our genomic alterations and molecular classification. Of the 55 autosomal genes uniquely overexpressed in primary glioblastoma, we found 13 to fall within regions that were preferentially amplified in primary versus secondary glioblastoma, compared with 4 that we would have expected based on genome size. Of the 21 autosomal genes uniquely overexpressed in secondary glioblastoma, 3 mapped to preferentially amplified regions whereas <1 is expected by chance. Taken together, these data suggest that a subset of these expression changes are driven by primary genomic events and that the many fewer genes found by Tso and et al. (2) to be differentially expressed may be due to the molecular heterogeneity of the disease that we observed within the genome. Finally, this study, along with that of Tso et al. (2), supports the view that the glioblastoma subtypes, indistinguishable on the histopathologic level, are driven by common and distinct genetic events.

Transformation to glioblastoma from low-grade diffuse astrocytoma in adults is considered to be an almost inevitable clinical outcome (48). Various treatment strategies have been implemented to delay transformation but, to date, no effective strategies have been identified owing mainly to the lack of understanding of this phenomenon and paucity of molecular targets. The identification of secondary glioblastoma-unique MCRs, as well as two clearly distinct genetic subgroups within the secondary glioblastoma tumors, provides an opportunity to investigate transformation to secondary glioblastoma. It is intriguing that there may be differences in clinical behavior, specifically time to transformation to glioblastoma in these subgroups (Fig. 4), although the lack of statistical significance of this finding underscores the need for an expanded sample set and an independent validation that these two classes differ by clinical progression time. Logistically, the number of candidate transformation loci is substantially reduced by subclassification of the secondary glioblastoma genome. Moreover, useful insights may derive from the identification of genomic events that track together in the subtypes. For example, the K3-3 subclass is interesting in its predominance of regional loss, suggesting that there are tumor suppressor genes embedded within the MCRs. Relevance to transformation to glioblastoma will ultimately benefit from detailed analysis of low-grade astrocytomas because genes important for tumor initiation would be present in both datasets whereas transformation genes would appear uniquely in the secondary glioblastoma. Along these lines, the chromosomal patterns within the subgroups might reflect the unique combinations of initiation and progression loci required for full transformation to glioblastoma, insights that could not be gleaned from the dataset without subclassification. How distinct genetic programs lead to a phenotypically indistinguishable end point in glioblastoma or how these genetic profiles will influence treatment responses remains a high priority for future investigation.

Acknowledgments

Received 6/8/2006; revised 9/22/2006; accepted 9/29/2006.

Grant support: The Goldhirsh Foundation (E.A. Maher and L. Chin); the Christopher S. Elliott Glioblastoma Brain Tumor Research Fund, The Richard Cerullo Fund, and Par Fore The Cure (E.A. Maher); NIH grants RO1CA99041 (L. Chin), CA57683 (D.N. Louis), and 5P01CA95616 (R.A. DePinho, D.N. Louis and L. Chin); and the Robert

A. and Renee E. Belfer Foundation Institute for Innovative Cancer Science (R.A. DePinho).

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.

We wish to thank Debra Gigas, R.N., Lisa Doherty, RN, ANP, OCN, Jennifer Zimmerman, and Louis Ostrowsky for assistance with case identification and clinical review.

References

- Kleihues P, Ohgaki H. Primary and secondary glioblastomas: from concept to clinical diagnosis. *Neuro-oncol* 1999;1:44-51.
- Tso CL, Freije WA, Day A, et al. Distinct transcription profiles of primary and secondary glioblastoma subgroups. *Cancer Res* 2006;66:159-67.
- Hui AB, Lo KW, Yin XL, Poon WS, Ng HK. Detection of multiple gene amplifications in glioblastoma multiforme using array-based comparative genomic hybridization. *Lab Invest* 2001;81:717-23.
- Nigro JM, Misra A, Zhang L, et al. Integrated array-comparative genomic hybridization and expression array profiles identify clinically relevant molecular subtypes of glioblastoma. *Cancer Res* 2005;65:1678-86.
- 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.
- Nishizaki T, Kubota H, Harada K, et al. Clinical evidence of distinct subgroups of astrocytic tumors defined by comparative genomic hybridization. *Hum Pathol* 2000;31:608-14.
- Quan AL, Barnett GH, Lee SY, et al. Epidermal growth factor receptor amplification does not have prognostic significance in patients with glioblastoma multiforme. *Int J Radiat Oncol Biol Phys* 2005;63:695-703.
- Rich JN, Hans C, Jones B, et al. Gene expression profiling and genetic markers in glioblastoma survival. *Cancer Res* 2005;65:4051-8.
- Burton EC, Lamborn KR, Feuerstein BG, et al. Genetic aberrations defined by comparative genomic hybridization distinguish long-term from typical survivors of glioblastoma. *Cancer Res* 2002;62:6205-10.
- Mellinghoff IK, Wang MY, Vivanco I, et al. Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors. *N Engl J Med* 2005;353:2012-24.
- 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.
- Schmidt MC, Antweiler S, Urban N, et al. Impact of genotype and morphology on the prognosis of glioblastoma. *J Neuropathol Exp Neurol* 2002;61:321-8.
- Tonon G, Wong KK, Maulik G, et al. High-resolution genomic profiles of human lung cancer. *Proc Natl Acad Sci U S A* 2005;102:9625-30.
- Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 2004;5:557-72.
- Carrasco DR, Tonon G, Huang Y, et al. High-resolution genomic profiles define distinct clinicopathogenetic subgroups of multiple myeloma patients. *Cancer Cell* 2006;9:313-25.
- Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* 2004;101:4164-9.
- 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.
- Brennan C, Zhang Y, Leo C, et al. High-resolution global profiling of genomic alterations with long oligonucleotide microarray. *Cancer Res* 2004;64:4744-8.
- Ohgaki H, Kleihues P. Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *J Neuropathol Exp Neurol* 2005;64:479-89.
- Misra A, Pellarin M, Nigro J, et al. Array comparative genomic hybridization identifies genetic subgroups in grade 4 human astrocytoma. *Clin Cancer Res* 2005;11:2907-18.
- Johansson FK, Brodd J, Eklof C, et al. Identification of candidate cancer-causing genes in mouse brain tumors by retroviral tagging. *Proc Natl Acad Sci U S A* 2004;101:11334-7.
- Ciafre SA, Galardi S, Mangiola A, et al. Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochem Biophys Res Commun* 2005;334:1351-8.
- Dunn M, Sinha P, Campbell R, et al. Co-expression of neuregulins 1, 2, 3 and 4 in human breast cancer. *J Pathol* 2004;203:672-80.
- Bieche I, Lerebours F, Tozlu S, Espie M, Marty M, Lidereau R. Molecular profiling of inflammatory breast cancer: identification of a poor-prognosis gene expression signature. *Clin Cancer Res* 2004;10:6789-95.
- Miller CT, Aggarwal S, Lin TK, et al. Amplification and overexpression of the dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2 (DYRK2) gene in esophageal and lung adenocarcinomas. *Cancer Res* 2003;63:4136-43.
- Dunant NM, Wisniewski D, Strife A, Clarkson B, Resh MD. The phosphatidylinositol polyphosphate 5-phosphatase SHIP1 associates with the dok1 phosphoprotein in bcr-abl transformed cells. *Cell Signal* 2000;12:317-26.
- Jung M, Romer A, Keyszer G, et al. mRNA expression of the five membrane-type matrix metalloproteinases MT1-MT5 in human prostatic cell lines and their down-regulation in human malignant prostatic tissue. *Prostate* 2003;55:89-98.
- Kuroyanagi H, Yan J, Seki N, et al. Human ULK1, a novel serine/threonine kinase related to UNC-51 kinase of *Caenorhabditis elegans*: cDNA cloning, expression, and chromosomal assignment. *Genomics* 1998;51:76-85.
- Toraman AD, Keser I, Luleci G, Tunali N, Gelen T. Comparative genomic hybridization in ganglioneuroblastomas. *Cancer Genet Cytogenet* 2002;132:36-40.
- Hughes S, Damato BE, Giddings I, Hiscott PS, Humphreys J, Houlston RS. Microarray comparative genomic hybridisation analysis of intraocular uveal melanomas identifies distinctive imbalances associated with loss of chromosome 3. *Br J Cancer* 2005;93:1191-6.
- Tong CY, Hui AB, Yin XL, et al. Detection of oncogene amplifications in medulloblastomas by comparative genomic hybridization and array-based comparative genomic hybridization. *J Neurosurg* 2004;100:187-93.
- Katoh Y, Katoh M. Identification and characterization of CDC50A, CDC50B and CDC50C genes *in silico*. *Oncol Rep* 2004;12:939-43.
- Belloni E, Trubia M, Mancini M, et al. A new complex rearrangement involving the ETV6, LOC115548, and MN1 genes in a case of acute myeloid leukemia. *Genes Chromosomes Cancer* 2004;41:272-7.
- Huang DP, Lo KW, Choi PH, et al. Loss of heterozygosity on the short arm of chromosome 3 in nasopharyngeal carcinoma. *Cancer Genet Cytogenet* 1991;54:91-9.
- Morohara K, Nakao K, Tajima Y, et al. Analysis by comparative genomic hybridization of gastric cancer with peritoneal dissemination and/or positive peritoneal cytology. *Cancer Genet Cytogenet* 2005;161:57-62.
- Bluteau O, Beaudoin JC, Pasturaud P, et al. Specific association between alcohol intake, high grade of differentiation and 4q34-q35 deletions in hepatocellular carcinomas identified by high resolution allelotyping. *Oncogene* 2002;21:1225-32.
- Chow SN, Lin MC, Shen J, Wang S, Jong YJ, Chien CH. Analysis of chromosome abnormalities by comparative genomic hybridization in malignant peripheral primitive neuroectodermal tumor of the ovary. *Gynecol Oncol* 2004;92:752-60.
- Natrajan R, Louhelainen J, Williams S, Laye J, Knowles MA. High-resolution deletion mapping of 15q13.2-q21.1 in transitional cell carcinoma of the bladder. *Cancer Res* 2003;63:7657-62.
- Gordon A, McManus A, Anderson J, et al. Chromosomal imbalances in pleomorphic rhabdomyosarcomas and identification of the alveolar rhabdomyosarcoma-associated PAX3-FOXO1A fusion gene in one case. *Cancer Genet Cytogenet* 2003;140:73-7.
- Wullich B, Sattler HP, Fischer U, Meese E. Two independent amplification events on chromosome 7 in glioma: amplification of the epidermal growth factor receptor gene and amplification of the oncogene MET. *Anticancer Res* 1994;14:577-9.
- Lin M, Smith LT, Smiraglia DJ, et al. DNA copy number gains in head and neck squamous cell carcinoma. *Oncogene* 2006;25:1424-33.
- Arai H, Ueno T, Tangoku A, et al. Detection of amplified oncogenes by genome DNA microarrays in human primary esophageal squamous cell carcinoma: comparison with conventional comparative genomic hybridization analysis. *Cancer Genet Cytogenet* 2003;146:16-21.
- Rossi MR, Conroy J, McQuaid D, Nowak NJ, Rutka JT, Cowell JK. Array CGH analysis of pediatric medulloblastomas. *Genes Chromosomes Cancer* 2006;45:290-303.
- Smeets SJ, Braakhuis BJ, Abbas S, et al. Genome-wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene-expressing human papillomavirus. *Oncogene* 2006;25:2558-64.
- Midorikawa Y, Tsutsumi S, Nishimura K, et al. Distinct chromosomal bias of gene expression signatures in the progression of hepatocellular carcinoma. *Cancer Res* 2004;64:7263-70.
- Meng S, Chen Z, Munoz-Antonia T, Wu J. Participation of both Gab1 and Gab2 in the activation of the ERK/MAPK pathway by epidermal growth factor. *Biochem J* 2005;391:143-51.
- Rich JN, Reardon DA, Peery T, et al. Phase II trial of gefitinib in recurrent glioblastoma. *J Clin Oncol* 2004;22:133-42.
- Maher EA, Furnari FB, Bachoo RM, et al. Malignant glioma: genetics and biology of a grave matter. *Genes Dev* 2001;15:1311-33.

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Marked Genomic Differences Characterize Primary and Secondary Glioblastoma Subtypes and Identify Two Distinct Molecular and Clinical Secondary Glioblastoma Entities

Elizabeth A. Maher, Cameron Brennan, Patrick Y. Wen, et al.

Cancer Res 2006;66:11502-11513. Published OnlineFirst November 17, 2006.

Updated version	Access the most recent version of this article at: doi: 10.1158/0008-5472.CAN-06-2072
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2006/12/01/0008-5472.CAN-06-2072.DC1

Cited articles	This article cites 48 articles, 17 of which you can access for free at: http://cancerres.aacrjournals.org/content/66/23/11502.full#ref-list-1
Citing articles	This article has been cited by 14 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/66/23/11502.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/66/23/11502 . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.