A SYSTEMS APPROACH TO THE DIFFERENT GRADES AND PROGRESSION OF HUMAN ASTROCYTOMA

BY

CHUNJING WANG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Chemical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Advisor:

Assistant Professor Nathan Price

ABSTRACT

Astrocytomas are the most common glioma, accounting for about half of all primary brain and spinal cord tumors. Malignancy in these tumors ranges from the least aggressive pilocytic astrocytoma (WHO grade 1) to the most aggressive glioblastoma (WHO grade 4). Molecular biomarkers or signatures—i.e., patterns of gene or protein expression that can reliably distinguish between each grade and provide insight into the underlying molecular events associated with tumor progression—have not yet been well established for astrocytomas. To identify candidate biomarkers and characterize genetic and molecular mechanisms driving glioma development and progression, we performed a meta-analysis of publicly available microarray gene expression datasets, comprising 432 tumor samples from all four grades and 28 non-tumor samples.

We first applied a consensus preprocessing method to raw microarray data to reduce bias introduced by different laboratories. Using DIRAC, a network-based classification approach previously developed in our lab, we were able to effectively differentiate tumor grades with an average accuracy of 87%. Additionally, we derived 46 specific transcriptional changes that are associated with astrocytoma progression; of the 46 genes, 27 were consistently upregulated and 19 were downregulated in the progression sequence.

Notably, we discovered a histology-independent classifier, a network using erythropoietin to mediate neuroprotection through NF-kB (EPONFKB), consisting of 11 genes and predictive of survival in high grade astrocytoma (HGA) patients. This network is known for its roles in neuronal development and is capable of classifying HGAs into previously unrecognized subtypes. It has proven to be a more significant survival predictor ($P = 2.4e^{-8}$) than histology-based grading ($P = 2.2e^{-6}$).

With our network signatures associated with each grade and our progression-associated genes, we hope to increase the understanding of molecular mechanisms leading to brain cancer development, maintenance and progression. With the identification of the EPONFKB network as a novel prognostic factor, we hope to move tumor diagnosis and prognosis toward a more quantitative realm.

ACKNOWLEDGEMENTS

The author wishes to express her appreciation to Dr. Nathan Price for his patient guidance in introducing the project to me, and for stimulating interesting discussions and overcome obstacles on the way to completion. She also owes gratitudes to James Eddy, who initially developed the DIRAC algorithm, which was heavily used in this project; Cory Funk, for helpful discussions on interpreting the results; Andrew Magis, for his help in uniform data processing; and Xianyang Zhang, for bringing in cold and refreshing drinks in this hot and humid summer. Last but not least, I want to express my deepest thanks to my parents and my sister who supported me not only in pursuing my master's degree, but also in all dreams in my life.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Astrocytomas are the most common gliomas, originating in star-shaped brain cells called astrocytes [1]. The conventional histopathologic diagnosis scheme is based on the World Health Organization (WHO) grading system, which assigns a grade from 1 (least aggressive, also called pilocytic astrocytomas) to 4 (most aggressive, also called glioblastoma) [2]. According to the WHO scale, major distinguishing hispathological features between different grades include growth rate of cells, rate of angiogenesis, and presence of necrosis [2].

Tumor typing and grading that rely completely on the WHO system may be insufficient due to the subjective nature of pathological diagnosis [3]. Histological variability is commonly present within the same tumor, whereby characteristics defined by WHO may only offer an oversimplified representation of the actual tumor features [3, 4]. As a result, diagnostic accuracy and reproducibility are jeopardized, giving rise to significant inter-observer variability [5, 6].

However, accurate diagnosis is required for adequate treatment and to assess prognosis for patients. While piloctyic tumors (G1) are easily removed through surgery, for low-grade tumors (G2) and higher, the chance of recurrence increases while the survival rate diminishes; anaplastic astrocytoma (G3) and glioblastoma (G4 or GBM) patients have dim prognostic prospects, with GBM few patients surviving more than 12 months after diagnosis [7, 8] .

Due to these reasons, many research studies have devoted their efforts to finding the genetic and molecular differences associated with astrocytoma patients. Most of these studies limited their focus to one or some of the grades [9-13]. Our study aimed to conduct a meta-analysis on microarray expression profiles of astrocytoma patients of various grades, and investigated the genetic and biological mechanisms implicated in the phenotypic differences. We achieved our goals through both network-based and gene-based approaches; firstly, networks that could best distinguish astrocytoma patients were analyzed followed by an investigation on the individual gene sets changing consistently with progression. Lastly, a network with prognostic value was found with available survival information.

A network-based classification method, Differential Rank Conservation (DIRAC) allowed us to identify differentially expressed networks (DENs) that revealed statistically robust differences between different astrocytoma tumors, leading to highly accurate classification of histologically similar phenotypes [14]. These molecular signatures represent the most perturbed networks in tumor samples as compared to non-tumor samples. It is known that tumorigenesis is a multi-step and complicated process; different biological networks associated with glioma evolvement become affected at different time points [15, 16]. By identifying these networks, we enhance our understanding of astrocytoma development from a network-based context.

Besides finding networks that could accurately predict histological grades of the patients, we also searched for networks that could predict survival. Among the four astrocytoma grades, high grades (HGA, include grade 3 and 4 tumors) have raised more interest than the other two grades, because of their sample availabilities and poor survival prospects [10, 11, 17]. Due to their molecular heterogeneity, these tumors could be classified further into smaller subtypes, either according to the path of progression (primary and secondary) or according to survival prospects (proneural, proliferative, mesenchymal). HGA usually occur *de novo* (primary), but may also progress from a lower grade (secondary) [18, 19]. Primary and secondary HGAs share similar morphologic features, and it has remained controversial whether they can be distinguished histologically [20]. Moreover, they are not clearly different in prognosis [19, 20]; this has led many to look for alternative subtypes showing survival differences. Philips *et. al.* classified HGA into either proneural (PN), mesenchymal or proliferative using a set of 35 genes. The PN subtype was shown to survive longer than non-PN subtypes [10, 11, 17].

Using unsupervised hierarchical clustering, we found a network, named EPONFKB, which could efficiently distinguish HGA patients into two groups with significant survival difference ($P = 2.4e^{-8}$), giving a much better separation between survival curves than when defined exclusively by histological grades ($P = 2.2e^{-6}$). The prognostic value of the subtypes outperforms that based on primary/secondary subclasses ($P = 0.001$) and is comparable to PN/non-PN subclasses ($P = 1.2e^{-8}$).

We defined the term “progression” following from the concept of secondary GBM, which has recurred in a patient with previously diagnosed lower grade astrocytoma. Since pilocytic tumors generally do not progress, the hypothetical sequence of progression is from normal to low-grade astrocytoma, progressing to anaplastic astrocytoma and eventually to GBM. The last part of our study was devoted to find progression-specific genes. From our datasets, two sets of genes that changed consistently in astrocytoma progression were identified and analyzed. One set consisted of 27 increasing genes and the other consisted of 19 decreasing genes, with both changing monotonically in the progressing sequence. These sets of genes allowed us to infer biological mechanisms as the tumor progresses. In doing so, we hope to provide genetic evidence associated with astrocytoma progression, which may help guide therapeutic decisions and eventually improve the clinical outcome.

CHAPTER 2

RESULTS/DISCUSSION

2.1 Effect of consensus preprocessing

Direct integration of datasets generated by different platforms, institutions and experiments introduces noise into data and may lead to the discovery of false biomarkers, therefore appropriate pre-computation is a necessary step in all meta-analysis studies [21]. Our pre-processing method greatly reduced sources of variations illustrated by an increase in Pearson correlation coefficients among patients. **Figure 1a)** is the Pearson correlation coefficient matrix obtained by applying GCRMA on raw microarray profiles of individual studies and combine them, and **Figure 1b)** was obtained by consensus pre-processing, which assembled raw expression data of 460 patients together and then applied necessary normalization steps (See **Materials and Methods**).

In **Figure 1**, laboratory effects are still prominent; due to the differences in sample characteristics, sample preparation, hybridization, and other protocol differences, tumor samples from the same institution looked more homogenous than samples obtained from different institutions. However, with consensus processing, these disturbances were greatly alleviated with the average correlation coefficient increased from 0.87 (**Figure 1**) to 0.95 (**Figure 2**).

Our processing method also allowed us to compute our confidence in whether a transcript is reliably present. We followed the probe sets detection algorithm implemented in Matlab Bioinformatics Toolbox (P-value for presence at 0.06, marginal at 0.04 and absence at P values below 0.04) and investigated how different probes filtering criteria (probes having 0%, or less than 25%, 50 %, 75% present calls for any phenotype could possibly be removed) affected the accuracy (**Figure A.1**) [22]. DIRAC showed robust performance against different numbers of probes present. Approximately 5000 probes that were absent throughout the whole phenotype were removed as this filtering strategy resulted in the best classification accuracy (**Figure A.1**). 15371 probes were kept for further analysis.

2.2 Global regulation of networks across phenotypes

Besides capturing how network expression patterns differ between phenotypes, DIRAC also provides quantitative measures (μ or rank conservation index) of how network rankings differ for a selected phenotype. If the combinatorial gene interactions in a specific network are quite similar among different patients, the network is considered *tightly regulated* within its phenotype. On the other hand, the network is considered *loosely regulated* if the ranks of network genes are greatly varied between samples of the same phenotype. Rank conservation index is a measure of the relative stability or consistency with which network rankings are maintained in a population [14].

Averaging rank conservation indices over all the networks provides a measure of global regulation in different phenotypes. For example, networks in normal patients samples are more highly conserved on average (0.962) than networks in pilocytic tumors (0.944), Similarly, network rankings in grade 2 to GBM samples matched the respective templates for 93.7% (μ= 0.937), 93.3% (μ=0.933) and 91.7% (μ=0.917) of all pairwise orderings on average. In fact, the relative magnitude of average rank conservation indices has an inverse correlation with the malignancy of phenotypes. This trend suggests that the more aggressive diseases may have greater overall variation in network ranking among different samples. In normal or low-grade astrocytoma patients, most networks are still under tight regulation to maintain normal cellular functions, as the tumor develops, more cooperating oncogenes involved in tumorigenesis and growth mechanisms were activated, consequently more biological networks become increasingly disturbed. This deregulation increases with grades may contribute to its malignancy. Based on a one-way ANOVA, the estimated overall P-value for the ordering of phenotypes in **Figure 3** is smaller than 0.001.

2.3 Top biomarkers from each classification

Instead of searching for best gene or gene pairs, our study aimed to search for a set of related genes that could accurately distinguish similar disease phenotypes. Expression levels of genes were grouped into 248 human signaling networks, defined according to the BioCarta gene sets collection in the Molecular Signatures Database (MSigDB) [23]. Microarray expression data of any two phenotypes among five disease/control states were combined and classified by DIRAC. Average cross-validation accuracy of 10 classifications is 86% (leave-one-out cross-validation) (**Figure 3)**. Two other manually curated network databases, Biology Process under Gene Ontology (BP), containing 825 curated pathways and cancer modules (CM) containing 456 modules (specifically related to cancer were also candidates for network database [23]. The average leave-one-out cross-validation accuracies of DIRAC on these two gene sets collections were 87.5% and 89.0% respectively (**Figure 3)**. Biocarta gene sets were chosen for further analysis based on two criteria: the relationships of the genes in each network are clearly defined with available interaction information, and the gene sets are small in size for time-efficient analysis.

Though different malignancies of astrocytomas are theoretically well defined by the WHO, drawing a distinction between them may be challenging based on histological grounds alone. Clinical and neuroradiologic features such as age, previous treatment often facilitate diagnosis and prognosis [17]; for instance, pilocytic astrocytomas are known for their high incidence in children and are associated with favorable prognoses [24], while infilatrating tumors of grade 2 to GBM occur usually in adults and are related to short survival [25]. However, these factors could be insufficiently decisive factors in diagnosis.

To select the best therapeutic decisions requires understanding of molecular alterations leading to tumor carcinogenesis and astrocytoma progression. Tissue microarrays have emerged as a popular tool for high throughput measurements of human genetic profiles, due to their ability to measure the tissue-specific protein expressions and identify possible treatment targets connected to the clinical outcome [26]. However, most of the microarray studies are focused on the high-grades, especially glioblastoma [9]. A meta-analysis including all malignancies of astrocytoma with normal brain tissues taken as control is necessary to gain insights into early and late events in the brain tumor evolution.

Malignant phenotypes, including cancers, usually arise as a net effect of interactions among multiple genes within networks. DIRAC is a powerful classification method based on the relative expression values of participating genes within biological networks. Furthermore, it identifies and measures network-level perturbations from a completely novel perspective, namely by the "combinatorial comparisons" of network genes as opposed to increases or decreases alone [14]. By accounting for these combinatorial interactions, we alleviated the signal-to-noise issues in disease-perturbed networks. DIRAC classified astrocytoma patients of different grades with each other and with normal samples with fairly high accuracy, with the majority of classifications between different phenotypes above 90% accurate except in cases among grade 2 to grade 4 tumors (G2 vs. G3 60%, G2 vs. G4 72%, G3 vs. G4 81%). This may be partially due to their invasive nature, which resulted in these tumors possessing histological and biological characteristics of more than one grade, posing challenges in the assignment of the grades; another reason is the small sample size of G2 tumors, which causes a less accurate training model and therefore, decreased precision in making predictions using DIRAC. In general, adjacent grades were similar in expression profiles and resulted in lower accuracies in these cases while far-apart grades were easier to distinguish.

The classification performance is also limited by other factors. One possible factor is the presence of the subtypes and/or outliers (atypical patients) within cancer tumor samples [11, 18]. Due to limited access to normal brain tissues, they were collected and defined in different ways and some degree of heterogeneity was introduced into the datasets inevitably. Other factors potentially affecting accuracies, e.g. the clinical variables such as ages and tumor locations of the patients, which should be controlled for a fair comparison, were not available in most cases.

Using a permutation-based testing to assess the statistical significance of estimated network classification rates (see Materials and Methods), we found a total of 219 and 195 networks which significantly discriminated between expression profiles of G1 versus normal controls and G2 versus normal controls respectively, while 211 and 217 networks were identified as robust signatures in distinguishing anaplastic astrocytoma and glioblastoma from controls ($P < 0.05$). Among these differentially expressed networks (DENs), we estimated that only 5~6% (between 11-12 networks in each set) are likely to have been found by chance rather than based on true

differences between the phenotypes, as determined by the FDR (see **Materials and Methods**). The top ten networks in each case were shown in **Table 1a)** to **1d)**

These networks represented significant changes that occurred in astrocytoma transitions. They could be functionally grouped into biological processes related to gliomagenesis and development such as cell cycle regulation, apoptosis signal transduction and immune system response [27].

2.4 Inferences from changes in network ranking

Different tumors vs. normal brains

Sixteen networks were more perturbed in early stage astrocytoma patients than in advanced stages as indicated by a decreased network ranking. Interestingly, these networks contain proto-oncogenes and are linked to early or initiation events in astrocytoma progression. An example is the platelet-derived growth factor (PDGF) signaling pathway whose ranks among all networks decreased with malignancy grade (**Table 1e**)**.** In response to various stress stimuli, the receptor tyrosine kinase PDGFR-α binds with its ligand PDGF-α, to stimulate various mitogenesis mechanisms [28]. The downstream molecules also in the PDGF signaling pathway include key players in the JAK-STAT and MAPK/ERK signaling pathways (JAK1, STAT1, STAT3, STAT5A and different isoforms of mitogen-activated protein (MAP) kinases, MAP2K1, MAP2K4, MAPK3, MAPK8). The PDGF signaling plays central roles during the initiation and progression of gliomas, and the overexpression of the receptors or the ligands has been found in both low/high astrocytomas [28-30]. The AKT pathway showed a similar pattern as the PDGF pathway in network ranking change, being more variably expressed in less malignant phenotypes (**Table 1e**); Phosphatidylinositol 3-kinases (PI3K) in the network converts phosphatidylinositol-4, 5-bisphosphate (PIP2) to the second-messenger molecule PIP3, which in turn phosphorylates **AKT** protein. Phosphorylated AKT triggers downstream pathways through activation of mammalian target of rapamycin (**mTOR**), transcription factor **NFk-β**, and MDM2 [31]. This network not only stimulates growth but also contributes to an increase in anti-apoptotic features of glioma cells [27]. PDGF and its associated activated downstream AKT signaling, acting in concerted efforts to promote cellular growth and survival explain these networks being more disturbed in early stage astrocytomas; however, while the tumor cells are active in growth and neuronal differentiation, they have not yet acquired angiogenic properties as in the high grades [32]. As the tumor progresses to more advanced stages, these networks become less disturbed probably due to the angiogenesis related networks being more activated

Six out of the sixteen networks are linked to immune system, one of the first responders in astrocytoma initiation. One such network, Cytokines and Inflammatory Response (INFLAM) showed similar expression pattern as the PDGF pathway (**Table 1e**). The network contains 8 families of interleukins (IL), as well as the tumor necrosis factor (TNF), and all three isoforms of tumor growth factor (TGFβ). Cytokines like IL-1 and IL-6 are involved in a broad range of

different immune networks to provoke the inflammatory response; TNF is able to induce apoptotic cell death, stimulate inflammation, and inhibit tumorigenesis [33]; TGF-beta suppresses proliferation and differentiation of lymphocytes including cytolytic T cells, natural killer cells and macrophages, thus preventing immune surveillance of the developing tumor [34]. Various immune mechanisms communicate and coordinate their efforts to prevent tumor growth. The network being highly differentially expressed in pilocytic astrocytomas but less variably expressed at later stages supported the theory that the immune system is alerted at a relatively early stage of gliomagenesis and could potentially restrain the tumor growth. At more advanced stages, an immune response may still be active, but may possibly be immediately overwhelmed by high tumor burden and fail to show a highly variably expressed pattern [35].

In contrast, DIRAC captured an opposite ranking pattern for thirteen other networks. One of which is Cell to cell adhesion signaling (CELL2CELL) are related to the angiogenic properties of HGAs (**Table 1e**). This pathway contains major cell adhesion proteins like Catenin, PECAM-1 and Paxillin. Catenins trigger changes in cell shape and motility; PECAM-1, involved in the formation of junctions between endothelial cells could modulate cell migration. Paxillin acts as an adaptor protein between proteins involved in adhesion signaling like FAK and SRC. These important molecules interact with cytoskeletal elements to produce changes in cell motility, migration, proliferation and shape [36]. Cell adhesion molecules have been associated with the invasive potential of GBMs, or a more aggressive subtype in GBM [11]. This network being more variably expressed in more malignant grades coincides with the fact that biological processes related to angiogenesis and cell invasion increase inactivity with aggressiveness of the tumor.

G1 vs other grades

Pilocytic astrocytomas represent a distinct pathological and biological entity compared with other tumors [24]. They behave as well-circumscribed tumors which do not diffusely invade the brain parenchyma; instead, the tumor could be removed by surgery and recurrent rate is fairly low. However, there are no studies which compared pilocytic astrocytoma patients with normal as well as all other grades from a network perspective. Therefore, we are interested in finding the networks differentiating pilocytic astrocytoma from higher grades and inferring biological events to explain its antimigratory properties. We combined microarray samples of malignancy grade 2 to grade 4 and termed this group collectively as diffusely infiltrating astrocytomas (DIA), indicating their nature to diffuse and spread. DIRAC detected 163 networks which could effectively separate grade 1 tumors from the rest ($P<0.05$) with their apparent accuracy ranging from 95% to 68%. The best network classifiers (with classification rate > 90%) are listed in **Table 1f)**.

Among the best five classifiers, four of them are related to cell cycle regulation (Mechanism of Gene Regulation by Peroxisome Proliferator-activated receptors (alpha) (PPAR), Cell Cycle: G1/S Check Point (G1), Cell Cycle: G2/M Checkpoint (G2), p53 Signaling Pathway (P53)). A lot of genes are shared among these networks implying extensive overlaps among, and cross-talk between, these pathways. PPAR agonists affect expression of cell cycle related proteins in cell lines of glial brain tumors; they decrease cell proliferation, stimulate apoptosis and induce morphological changes and expression of markers associated with better prognosis [37]. P53 is a guardian in G1/S phase, whose inactivation allows cell cycle progression and makes apoptosis mechanisms ineffective. This tumor suppressor was shown to be not involved in the oncogenesis of pilocytic astrocytomas [24]; but for diffuse astrocytomas of WHO grade II, frequent mutation (occurs in up to 82%) of TP53, which encodes the protein was detected and believed to be one of its defining characteristics [16]. These networks also contain proto-oncogenes MDM2, which is a negative regulator of p53, members of the cyclin-dependent kinase (CDK) family (CDK2, CDK4 etc.), and their inhibitors (CDKN 1A, CDKN 2A etc.). Similar to TP53, homozygous deletion of CDKN2A is found of 20% in diffuse astrocytoma while remains intact in grade 1 tumors [24], indicating these changes represent genetic events occurred to grade 2 or higher grade tumors. Altered combinatorial interactions among gene pairs in the cell cycle regulation networks in

pilocytic astrocytomas as compared to DIAs may contribute to effective tumor suppression mechanisms such as cell-cycle arrest, which eventually results in controlled tumor cell growth and less tendency to migrate to brain parenchyma than other grades.

2.5 Monotonically changing genes in astrocytoma progression

Certain genes appeared to be repeatedly differentially expressed as we compared different tumor grades against controls; this led us to search for a list of genes whose expression level were consistently up or down regulated in human astrocytoma progression. In doing so, we hope to correlate the molecular changes with the histopathological development. The commonly recognized path of progression on a normal person starts with the onset of low-grade astrocytoma, skipping the first grade, progressing to anaplastic astrocytomas and eventually to GBM, though patients may not necessarily go through each stage. However, this model could hardly be proven clinically due to the lack of available tissues in the early phases of tumorigenesis. In this aspect, spontaneous genetically engineered mouse models provide an opportunity to track the molecular and pathological changes as a function of time. Low-grade astrocytoma started to develop as early as from 1-week old genetically engineered mouse; from 3 to 8 weeks the incidence of low-grade astrocytomas progressively increased, with 85% of 12-week-old mice harboring low or high-grade astorcytomas [38].

"Progression" in human astrocytomas was established as mentioned before, with the support of mouse models. We kept track of differentially expressed genes from normal to GBM and found 27 genes that successively increased and 19 genes that similarly decreased their relative expression values. Considering both our increasing and decreasing genes, we grouped them into four functional categories according to in DNA damage repair, chromatin regulation and apoptosis (**Table A.1-2**), according to the putative functions they encode.

DNA Damage Repair related genes

Our list of successively changing genes includes several genes that play critical roles in specific DNA damage repair processes. Those genes include DCLRE1B, PALB2, RBBP8, TOP1 and SMARCA2; all of these, with the exception of SMARCA2, increased their relative expressions in astrocytoma progression (**Table A.1-2**). The susceptibility to DNA damage arises from a compromised repair system, either in the repair proteins themselves or in the DNA damage response pathways. Genomic instability and the susceptibility to DNA damage can be indicative of a poor prognosis in many cancers, including gliomas.

DCLRE1B, also known as Apollo, is a 5’ to 3’ exonuclease that helps to maintain the 3’ overhang of telomeres, thereby suppressing activation of the DNA damage sensing protein [39]. Along with its binding partner TRF2, Apollo was also reported to help relieving topological stress during telomere replication in S phase and protecting chromosome termini from being recognized and processed as DNA damage [40, 41]. Increased levels of Apollo thus may repress the activation of DNA-damage or apoptosis mechanisms in astrocytoma progression.

Apollo negatively regulates Topoisomerase 1 (TOP1) [42], another upregulating gene that controls and overcomes topological problems during DNA transcription. It was implicated in all three functional classes, but it has been mostly associated with DNA damage repair [43-45]. Inhibitors of TOP1 have shown promising activity in patients with high grade gliomas and warrant further study [46].

TOP1 was also regulated by RBBP8 in our list, commonly known as CtIP, which functions in DNA double strand break repair by homologous recombination [42, 47, 48]. It has also been characterized as a transcriptional cofactor, interacting with and modulating the activity of the transcriptional repressor CtBP [49], the DNA repair protein BRCA1[50, 51], and G1/S-specific protein CyclinD1 as well as the oncogene retinoblastoma (Rb) [52, 53]. Due to its association with BRCA1, it had been heavily studied in breast cancer, with the corresponding protein expression considered as useful biomarkers for breast cancer prognosis; in addition, CtIP silencing was reported to be a novel mechanism for effective breast cancer therapy [54]. Another increasing gene, PALB2, last one in this category, enables recombinational DNA repair, in ways similar to RBBP8, through binding with BRCA2 [55]. Its increase in activity may imply its roles in cell cycle progression leading to tumor cell proliferation, in addition to preventing apoptosis of tumor cells in astrocytoma progression.

Chromatin remodeling related genes

All of the human genome is packaged into chromatin, which is continuously remodeled. The fate of the cell relies on a delicate balance between gene expression and repression. The transcriptional control of the genome is maintained not only by transcription factors but also chromatin remodeling proteins [56]. 4 increasing genes (PRMT1, ENY2, ESPL1, NCAPG

WDR76) and one decreasing gene SMARCA2 identified in our list, displayed significant connections to the chromatin modification process, as listed in **Table A.1-2**.

The building blocks of chromatin, named the nucleosome, can be restructured by two mechanisms: 1. the movement of nucleosomes by ATP-dependent chromatin remodeling complexes; and 2. the modification of core histones by histone acetyltransferases, deactylases, methyltrans-ferases, and kinases [56, 57]. Our gene set modified the chromatin through both mechanisms.

As an arginine-specific histone methyltransferase, PRMT5 is primarily known to be involved in epigenetic gene regulation through histone methylation of H2A, H3 and H4 [58, 59]. It also has the ability to facilitate dedifferentiation and to create pluripotent stem cells, resembles functions of biomarkers indicative of mesenchymal high grade glioma (HGA), a more malignant subtype associates with worse prognosis.

PRMT5 also methylates several other important glioma-related genes including p53, JAK2 and EGFR. By interacting with PRMT, p53, a tumor suppressor, was weakened in its target specificity and ability to facilitate DNA damage repair [60-62]; PRMT5 has also been linked to the invasive and migratory potential of glioblastoma cells through its regulation of the ubiquitously expressed tyrosine kinase JAK2 [63]. Further extending PRMT5's role in oncogenesis is the observation that activation of the JAK2/STAT3 pathway correlates with glioma grade and aggressiveness, and that this activation occurs more frequently in gliomas expressed with EFGR, another reported PRMT5 methylation target [64]. Taken together, increased expression of PRMT5 may directly or indirectly contribute to a cascade of events that lead to the progression of early-grade tumors to later more malignant phenotypes.

SMARCA2, better known as *Brm*, is a ATP-dependent chromatin complex. Consistent with our observation that *Brm* monotonically decreased with respect to glioma grade, aberrant expression of *Brm* genes is associated with disease development and progression in many cancers [65, 66]. Though the exact mechanism is not clear, there has been speculation that the loss of *Brm* may result in the increased interactions with transcription factors such as Oct4 and Sox2 to enable pluripotency, which leads to tumor cell proliferation and [67-69]. Lastly, *Brm* is also known to

interact with PRMT5 resulting in further chromatin remodeling through the methylating action of PRMT5 on H3 and H4 [70].

The remaining two genes ENY2 and ESPL1, were also implicated in regulation of gene expression. ENY2 has been identified as a part of another chromatin remodeling complex, SAGA, known for its ability to acetylate histones H2A and H2B [71]; On the other hand, ESPL1 plays a pivotal role in the separation of sister chromatids at anaphase. Overexpression and mislocalization of ESPL1 were seen in a wide range of human cancers [72] .

Apoptosis

Many genes in our list encoded apoptotic proteins. Unlike the other two functional categories, the majority of apoptotic genes (6 out of 10) decreased with tumor grade (**Table A.1-2**). Because evasion of apoptosis is generally considered as a hallmark of cancer, pro-apoptotic genes that suppress formation of tumors should decrease their expression with tumor grade, for instance C10orf97 and DSTYK. In contrast, anti-apoptotic genes that facilitate tumorigenesis should be more active in high-grade gliomas, as in the case of NUP107. However, we also found AK2 and BCL2L11 who seemed to induce cell death, became more active as the tumor progressed; to support our finding. a number of studies validated their overexpression in various diseases, including cancer,

One possible explanation to this interesting phenomenon is that genes and their respective encoded proteins usually play a number of roles in signal transduction, and apoptosis may be one of the many functions known to us, therefore the up or down regulation of a certain transcript is a net effect of biological pathways interacting and influencing each other rather. We will need more discoveries regarding their exact roles in cancer, or astrocytoma in particular. Nevertheless, a large number of genes related to programmed cell death, allowed us to further understand its involvement in glioma development and progression.

Blood biomarkers

Four of the key transcripts (ESPL1, KIF15, NUP205, PRMT5) in this list encoded proteins that are secreted into the blood. These blood biomarkers have significant potential applications in early detection and management of various diseases, including cancer. Simply by measuring

altered blood levels of proteins without the need to sample disease tissues is a less expensive and more friendly therapeutic option for disease prediction and monitoring than traditional medicine.

2.6 Prognostic networks that reclassifies high grade astrocytoma

HGAs are very aggressive tumors usually associated with worst prognosis, with average length of survival for G3 patients around three years, and fourteen months for GBM patients [73]. Due to the heterogeneity of their molecular profiles, a lot of efforts had been devoted to discover subtypes within the large collection of tumor samples. A long-existing attempt to classify HGA is according to the clinical histories of patients: a HGA tumor may be classified as either primary if there was no prior history of tumor occurrence, or secondary, if the tumor recurred to the patients, usually in a more malignant form. Primary tumors are believed to occur to older patients, and have a slightly shorter survival time [18].

Another relatively new approach by Phillips *et al*, utilized a set of 35 genes to classify HGAs into three classes, and resembled each subclass with a corresponding stage in neurogenesis . One tumor class (proneural or PN) displaying neuronal lineage markers shows longer survival, while two other tumor classes enriched for neural stem cell markers display equally short survival (non-proneural or non-PN).

Our study was also interested in finding biomarkers indicative of HGA prognosis. With available survival information from 239 patients, we derived a novel network with significant prognostic value. Using a distance matrix defined based on the DIRAC metric (see **Materials and Methods**), the genomic profiles of all HGA patients were grouped into two clusters using unsupervised clustering. Subsequent log-rank tests on the survival estimates of these two groups gave a P-value of $2.4e^{-8}$, comparable to the performance given by the genes separating proneural vs, non-proneural groups (P value=$1.2e^{-8}$); more importantly, it outperforms p-values given by histological separation (grade 3 against grade 4 tumors, P value=0.001) as well as by path of progression (primary against secondary tumors, P value=$2.2e^{-6}$). Besides this network, four others defined in the Biocarta network collection yielded statistically significant P values in their respective log-rank tests (**Table A.3**)

The EPONFKB network utilizes erythropoietin to mediate neuroprotection through NF-kB. It initiates signaling when erythropoietin (EPO) binds with its receptor EPOR to trigger NF-kappa-B, a heterodimeric complex formed by the Rel-like domain-containing proteins RELA/p65 and NFKB1/p50 in the presence of Janus kinase 2 **(**JAK2). The activated complex then translocates

to the nucleus and subsequently sends out both apoptotic signals through superoxide dismutase 2 (SOD2) and antiapoptotic signals through glutamate receptor subunit zeta-1 (GRIN1).

Involvement of the EPONFKB network in gliomas is an interesting and debatable subject as how it affects survival. The standard treatment of GBM with radiation does significant damage to the surrounding brain, resulting in significant collateral damage. This damage is often referred to as "radiochemobrain" and results in slowing psychomotor skills, cognitive decline, fatigue, and loss of drive, all of which significantly reduced quality of life [74, 75]. To counteract these effects, patients have been given hematopoetic growth factor erythropoietin prior to and following radiation. EPO signaling crossactivates the antiapoptotic transcription factor NF-kappaB, this causes neuroprotection against oxidative stress and implies radioprotection. As a matter of fact, EPO has pleotrophic affects on the brain including anti-apoptotic, antioxidative, neurotrophic, axon-protective, angiogenic, and neurogenic – many of which appear to be neuroprotective for the insults of radiation and chemotherapy [76-82]. In addition, EPO has also been shown to improve the responsiveness of tumors to radiation therapy in human glioma xenographs by increasing tumor oxygenation [83, 84].

Despite all the positive effects EPO it believed to have, these same effects could also help to promote tumor growth. Administering EPO as part of chemotherapy is controversial because of the possibility for it to promote tumor growth [85-88]. Recently, EPO signaling was shown to be involved in angiogenesis of human glioma cells as well as cancer stem cell maintenance [89, 90]. Still, others have shown that while EPO does augment the survival of glioma cells, it is unlikely to appreciably influence basal glioma growth [91].

While our results implicate the EPONFKB as a novel and powerful biomarker in predicting patient survival, the exact mechanism through it modulates survival is unclear. One possible explanation is the longer survival subtype with 161 patients may have more patients underwent EPO-related chemotherapy; this group also has a larger proportion of patients (45%) belonged to the PN subtype, as compared to the shorter survival subtype (13%), consistent with the observation that PN class has better prognosis than non-PN classes.

2.7 Tables

Network Name	Apparent Accuracy	*P*-values
G2	0.984	<0.001
G1	0.984	<0.001
TID	0.977	<0.001
PDGF	0.977	<0.001
EGF	0.977	<0.001
IL1R	0.977	<0.001
P38MAPK	0.977	<0.001
GSK3	0.974	<0.001
HIVNEF	0.974	<0.001
EIF4	0.973	<0.001

Table 1. Top 10 networks differentiating normal from grade 1 patients

Network Name	Apparent Accuracy	*P*-values
PROTEASOME	0.982	<0.001
EGF	0.982	<0.001
MCALPAIN	0.964	<0.001
ALK	0.964	<0.001
CBL	0.949	<0.001
GSK3	0.949	<0.001
PDGF	0.948	<0.001
AT1R	0.946	<0.001
EIF4	0.941	<0.001
FAS	0.940	<0.001

Table 2. Top 10 networks differentiating normal from grade 2 patients

Network Name	Apparent Accuracy	*P*-values
ERK	0.991	<0.001
CHEMICAL	0.955	<0.001
EGF	0.954	<0.001
MCALPAIN	0.954	<0.001
KERATINOCYTE	0.945	<0.001
IGF1MTOR	0.944	<0.001
CBL	0.935	<0.001
BCR	0.935	<0.001
GSK3	0.935	<0.001
IL2RB	0.934	<0.001

Table 3. Top 10 networks differentiating normal from grade 3 patients

Network Name	Apparent Accuracy	*P*-values
CELL2CELL	0.962	<0.001
P38MAPK	0.950	<0.001
G2	0.949	<0.001
G1	0.948	<0.001
ERK	0.947	<0.001
MPR	0.943	<0.001
PROTEASOME	0.936	<0.001
VEGF	0.935	<0.001
CELLCYCLE	0.934	<0.001
HIVNEF	0.934	<0.001

Table 4. Top 10 networks differentiating normal from grade 4 patients

	G1			G2			G3			G4		
	Ranking	Classification Rate	FDR	Ranking	Classification Rate	FDR	Ranking	Classification Rate	FDR	Ranking	Classification Rate	FDR
PDGF	4	0.977	0	7	0.948	0	32	0.899	0	66	0.894	0
AKT	65	0.939	0	122	0.840	0.0008	141	0.810	0.00001	151	0.824	0
INFLAM	73	0.935	0	99	0.862	0.002	107	0.851	0	155	0.821	0
CELL2CELL	116	0.908	0	81	0.871	0.002	65	0.873	0	1	0.962	0

Table 5. Changes in network rankings of selected networks in different astrocytoma grades vs. normal controls; PDGF, AKT and INFLAM showed decreased ranking as the diseases develop, while CELL2CELL showed an opposite pattern

Network Name	Apparent Accuracy	FDR
PPARA	0.946	0
G2	0.932	0
KERATINOCYTE	0.926	0
P53	0.915	0
G1	0.911	0

Table 6. Top 5 networks differentiating grade 1 from other astrocytoma tumors

Platform	Authors of Study (year, GSE accession)	Number of patients in each class				
		Normal	Grade 1	Grade 2	Grade 3	Grade 4
U133A	Frejie *et al* (2006, GSE 4412) [10]	0	0	0	8	46
	Phillips *et al* (2006, GSE 4271) [11]	0	0	0	21	55
	Wong *et al* (2008, GSE 12907) [12]	4[a]	21	0	0	0
	Rich *et al* (2005, GSE 13041) [38]	0	0	0	0	31
	Lee *et al* (2008, GSE 13041) [17]	0	0	0	0	28
	Barrow *et al* (2008, GSE 13041) [17]	0	0	0	0	31
	Mcdonald *et al* (2005, GSE 3185)	0	0	3	0	0
Total U133A		4	21	3	29	191
U133-Plus 2.0	Sun *et al* (2006, GSE 4290) [13]	23[b]	0	7	19	77
	Liu *et al* (2010, GSE 19728)	1[c]	2	5	5	5
	Sharma *et al* (2007, GSE5675) [39]	0	41	0	0	0
	Lee *et al* (2008, GSE 13041) [17]	0	0	0	0	27
Total U133-Plus 2.0		24	43	12	24	109
Total		***28***	***64***	***15***	***53***	***300***

Table 7. **Summary of microarray expression datasets included in the study**

[a]One normal fetal brain RNA, 1 normal cerebellum RNA, 2 normal tissue were surgically removed tissue adjacent to resected tumor tissue and RNA extracted

[b] Brain samples of epilepsy patients

[c] Pooled normal brain tissue

2.8 Figures

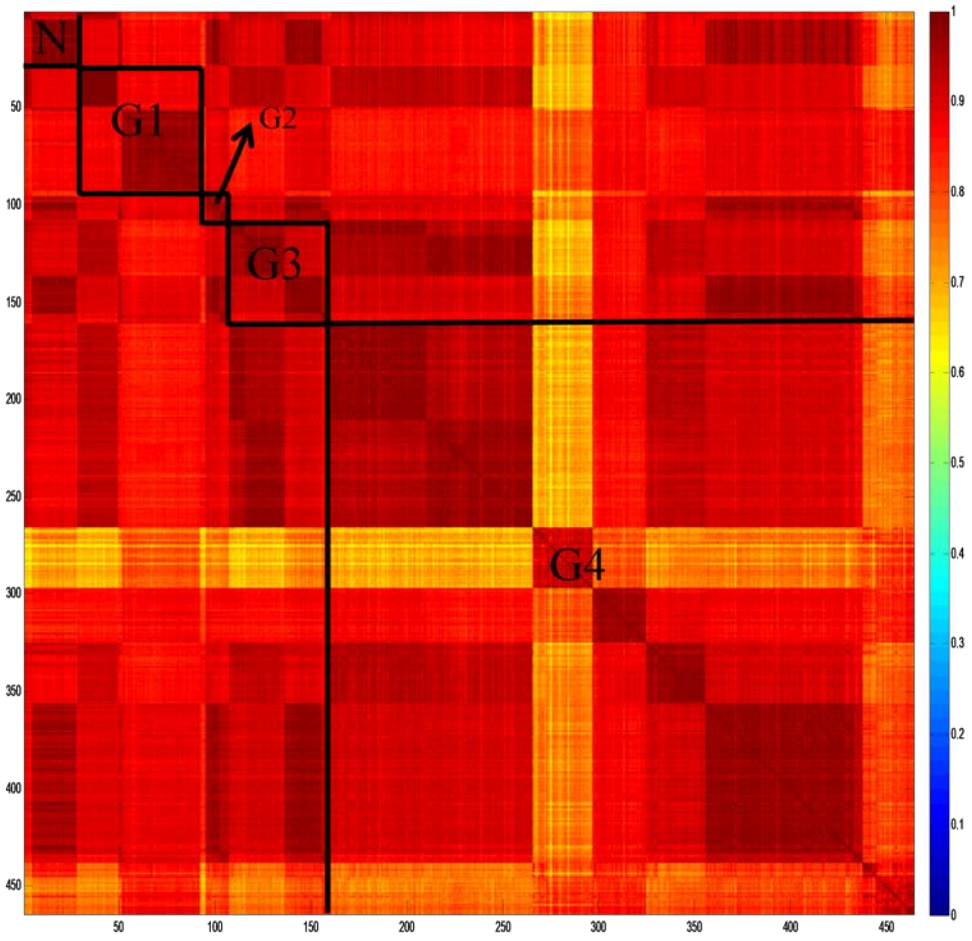


Figure 1. Pearson-correlation matrix before consensus pre-processing. Tumor samples from the same experiments displayed higher homogeneity than other samples

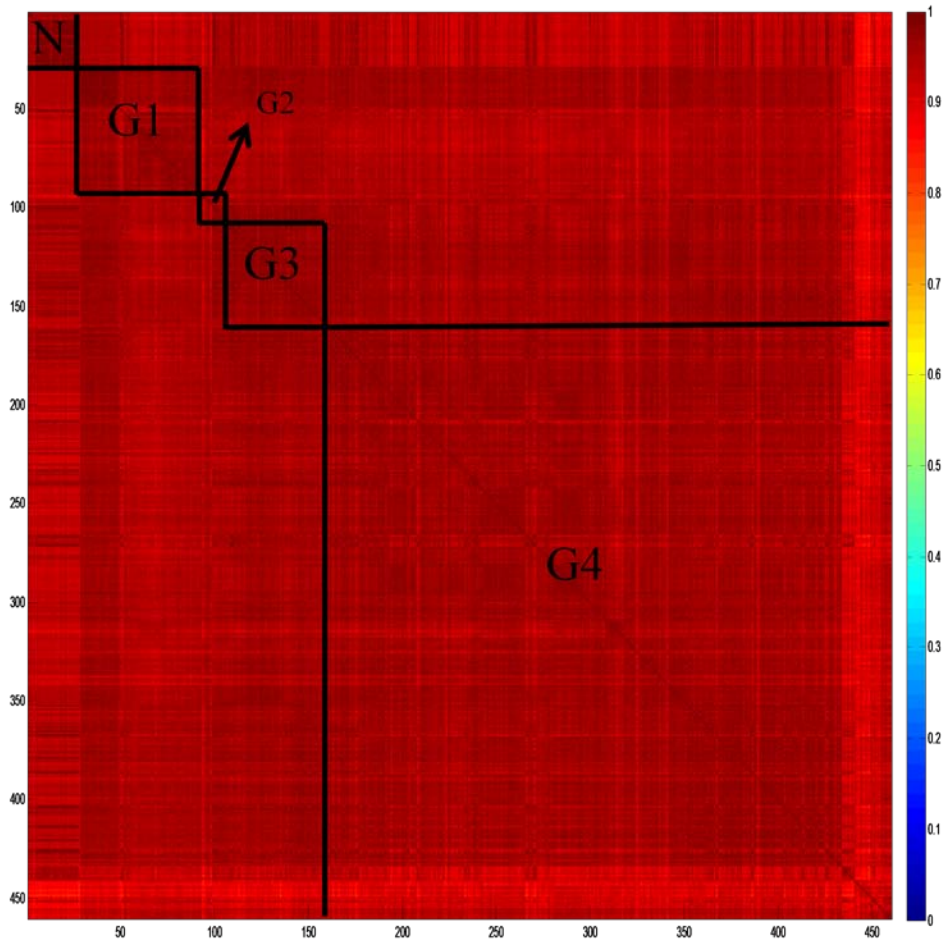
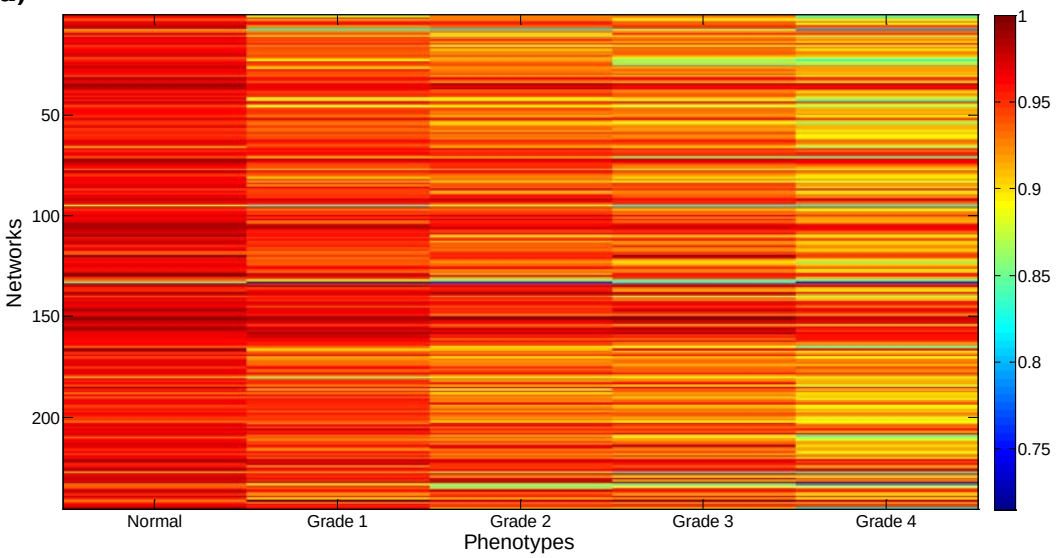


Figure 2: Pearson-correlation after consensus pre-processing; Laboratory effects are much less obvious; tumor samples from different studies or phenotypes all look highly correlated.

3 a)



3 b)

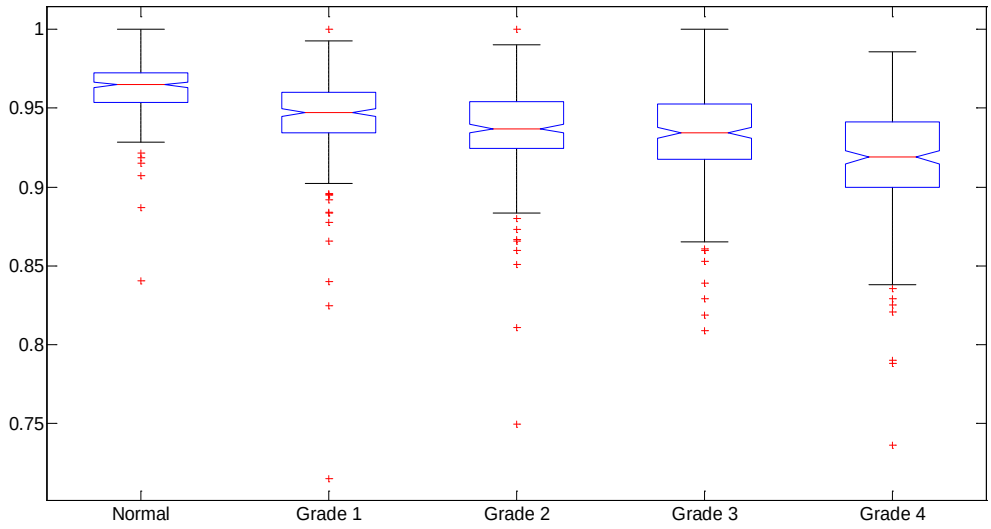


Figure 3: **Rank conservation of networks across phenotypes**. Colors on the heatmap represent rank conservation indices for each network in the five different phenotypes, where the brighter colors indicate very tight regulation of network ranking in a phenotype and the darker colors indicate loose regulation of networks.

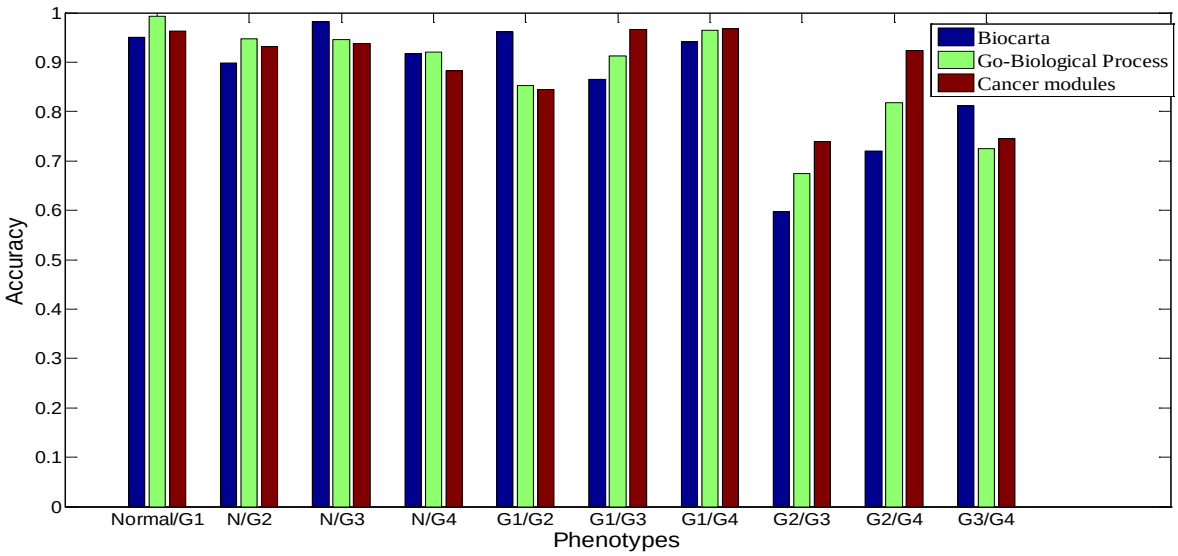
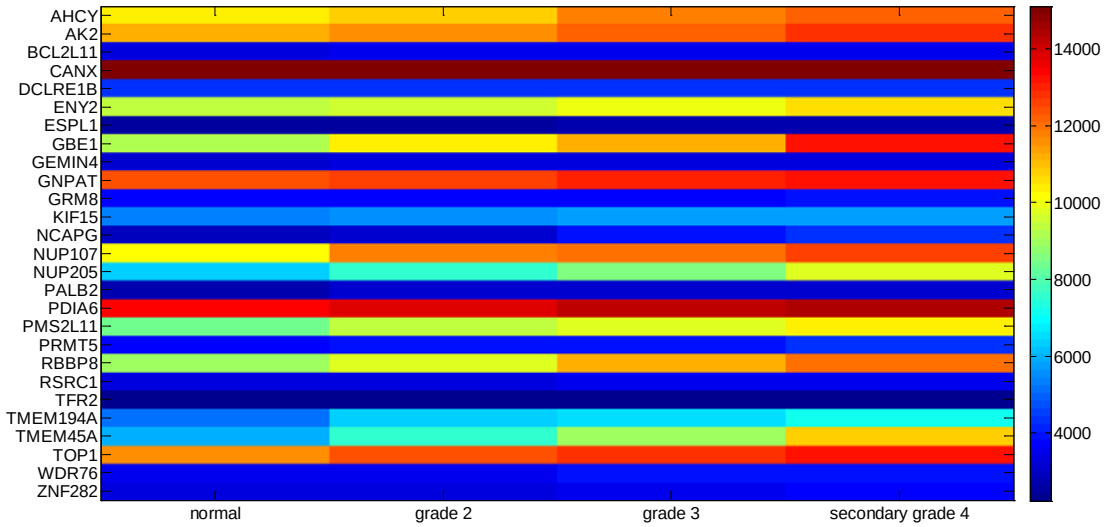


Figure 4: **Classification accuracy on different network databases**; cross-validation accuracies of DIRAC using three different network collections are fairly close

5 a)



5 b)

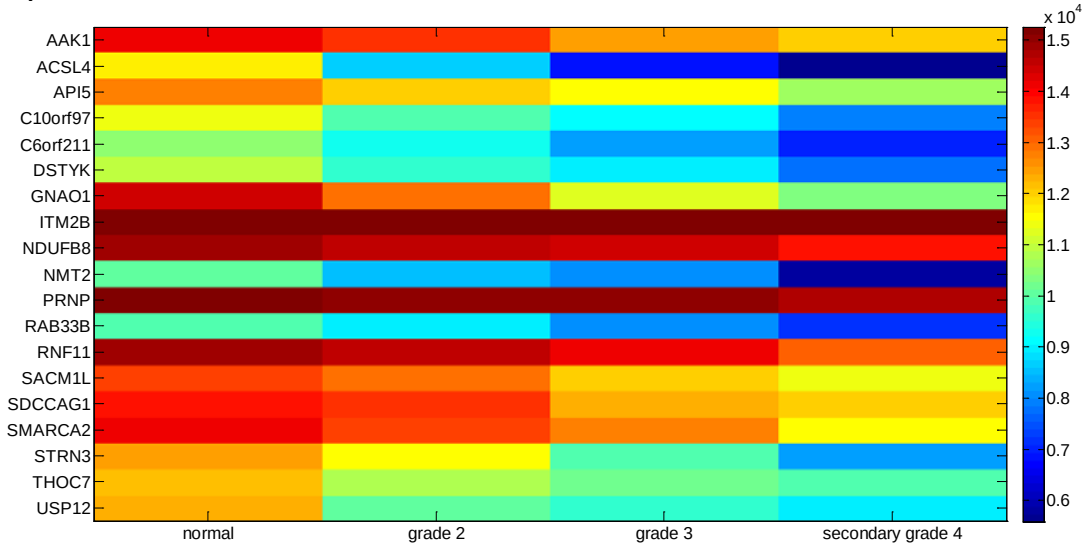


Figure 5: Genes showing consistent dys-regulation in progression a) 27 upregulated genes b) 19 downregulated genes

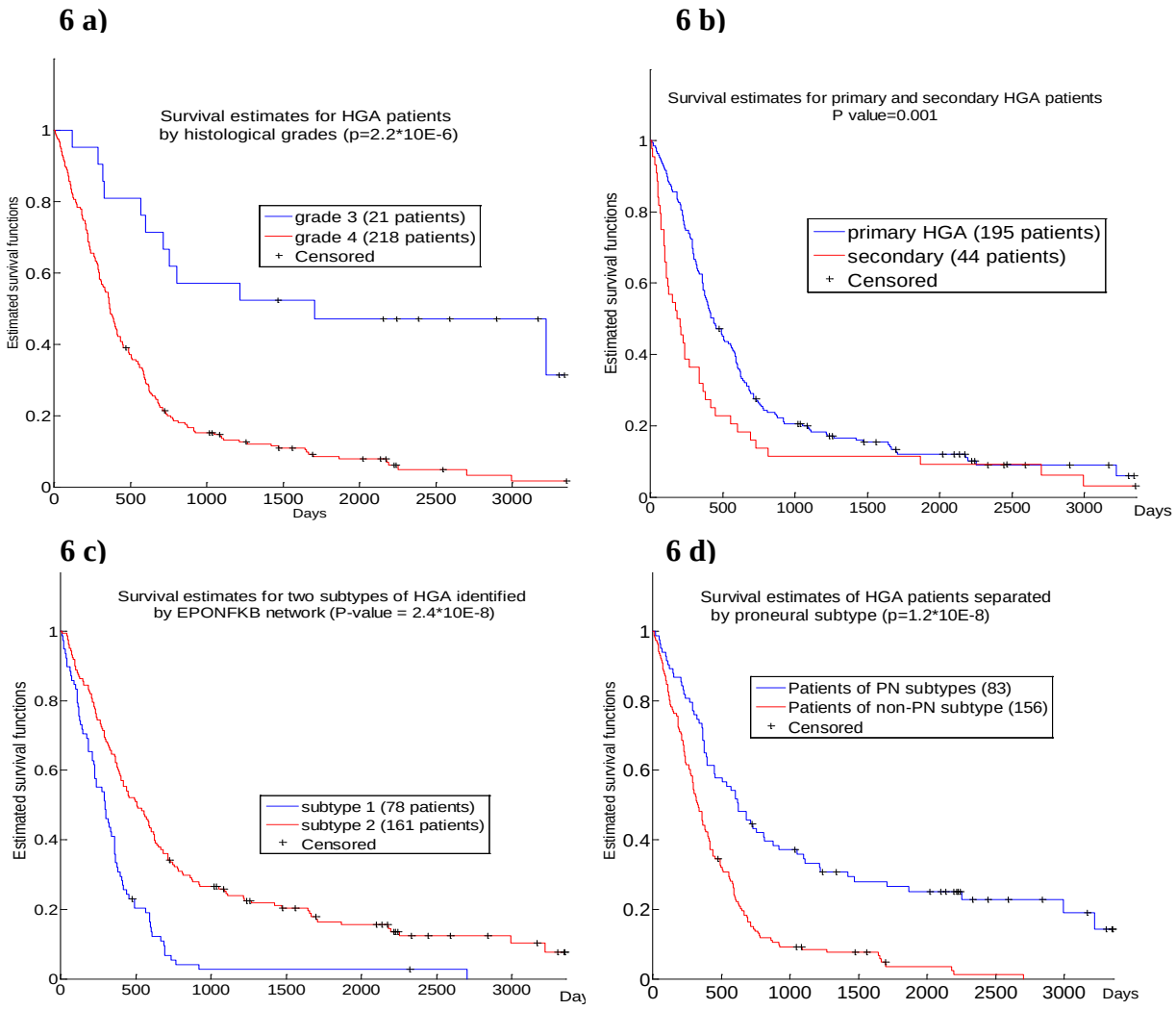


Figure 6: Comparison of different approaches to re-classify HGAs. a) Survival estimates of two groups separated by histological grades b) by primary or secondary HGA subtype c) by prognostic network marker EPONFKB d) by proneural (PN) or non-PN subtype. The log-rank test on the two subtypes defined by our network marker has a p-value comparable to one of the best regarding scheme.

CHAPTER 3

CONCLUSIONS

DIRAC is a classification approach accounted for the combinatorial behavior of interacting genes in biological networks, which provides accurate molecular signatures between different astrocytoma grades. These signatures allowed us to learn the most disturbed networks associated with tumor malignancy. Moreover, a network consisted of 11 individual genes was found to be significant in predicting prognosis of high grade astrocytomas. Besides network biomarkers, 46 genes were observed to change their expression values in a unidirectional manner with the tumor grade; their biological functions and implications in astrocytoma progression were identified.

The significantly perturbed pathways identified in this study offered biological insights into tumorigenesis and progression of astrocytoma, and may be potential candidates for novel diagnostic approaches. Due to the high heterogeneity of malignant brain tumors, the ideal and most effective therapy should be treatment personalized for each individual patient. Clinical therapies could be developed based on assessing the differentially expressed signaling networks, and targeting of specific network alone or in combination with traditional therapy.

CHAPTER 4

MATERIALS AND METHODS

4.1 Collection of microarray data

Raw microarray CEL files (Affymetrix, Santa Clara, CA) were retrieved from previous studies deposited to the publicly accessible microarray database, the NCBI Gene Expression Omnibus (GEO) between year 2005 and 2010. The data selection criterion is that the RNA of the tumor samples had to be hybridized to either one of the two platforms, Affymatrix HG-U133A or its complimentary version, HG-U133-plus 2 GeneChips. Affymatrix HG-U133B was not chosen since it does not include any probes common with U133A; this mutual exclusiveness of probes would add heterogeneity into the data, which would increase noise in separating different phenotypes. The GEO accession number, year of publication, together with number and grades of samples reported in each original study are listed in **Table 7**.

4.2 Integration of microarray data

A unique "consensus preprocessing" method was used on microarray CEL files to normalize differences introduced by different studies and individual preprocessing methods. Common probes (22277 probes) shared by two microarray chips (U133A and U133-Plus 2.0) were identified according to publicly available array descriptions, followed by GeneChip RMA (GCRMA) with the default settings from Matlab (MATLAB version 7.7, The MathWorks Inc., 2008). GCRMA was selected because its performance was observed to be superior to RMA and other normalizations for background adjustment of multi-arrays on the unified gene expression matrix [22].

In order to convert the probe intensity matrix to a gene expression matrix, the probe set lists were annotated according to definitions of these Affymatrix GeneChips. Probes mapping to multiple genes were removed from the probe lists; in cases where multiple probes correspond to the same gene, the maximum intensity was computed to determine the corresponding expression value for that gene. In the last step of preprocessing, all absolute intensity values were replaced by their relative ranks within each array.

4.3 Classification using DIRAC

Networks consist of components (nodes) and interactions (edges) between them. The nodes could be metabolites and macromolecules such as proteins, RNA molecules and gene sequences, while the edges are physical, biochemical and functional interactions. Our study considered three manually curated gene sets (BIOCARTA, Cancer Module and GO-BP) collected from various sources such as online pathway databases, publications in PubMed, and knowledge of domain experts [40].

For each selected network, DIRAC computes the expected ranking of network genes (rank template) by averaging the individual ordered gene expression profile within each phenotype, and measures how closely each sample's network ordering matches the phenotype-specific template (rank matching score). Class labels were assigned based on similarity of the patient's individual profile to one particular template of the two templates, and apparent accuracy was calculated based on percentage of correct prediction of phenotypes; the estimated classification accuracies for all networks were calculated likewise [14]. A null-distribution of network classification rates were generated by randomly reassigning the original phenotype labels 1000 times. A significance level was measured as the probability of observing classification rates in the null distribution greater than or equal to the real rates. To address the issue of multiple-hypothesis testing, the significance level was adjusted to false discovery rate (FDR), representing the fraction of expected false positives [14].

We used leave-one-out cross validation to estimate the error rate of DIRAC. Importantly, all processes including defining rank templates, and selecting the best network were done within cross-validation, using only the training samples (i.e., no information from test samples were used to train classifiers) [14]. The overall cross validation classification rate was calculated from the average of sensitivity and specificity of all predictions in each set.

4.4 Rank conservation indices

The rank template for each phenotype and rank matching scores of each patient to the template were calculated as explained earlier (for details, please refer to [14]). The rank conservation index is the average of the all matching scores of samples within a certain phenotype. By averaging the rank conservation indices of all networks for a disease, we had a single value measuring the relative deregulation of the networks for that phenotype.

4.5 Monotonically increasing/decreasing genes

In each adjacent pair of astrocytoma grades in the progression sequence, differentially expressed genes (DEGs) were selected based on the Wilcoxon ranksum test ($P < 0.05$ after Bonferroni correction). The intersection of these gene sets represented consistently deregulated genes. 27 DEGs showed consistently positive $\log_2$ expression ratios in each gene set and thus are the monotonically increasing genes in progression, while another 19 decreased their expression as the disease progresses.

4.6 Subtyping HGA

Time of survival (days or weeks) and subtype designations were available for 239 patients. The microarray expression matrix of this subset of patients was normalized as previously described. A distance matrix was constructed for each selected network based on the pairwise orderings of the genes within the network. For example, if a network *P* consisted of six genes, there could be $\binom{6}{2} = 15$ distinct ordered pairs. For a gene pair i and j, let X denote their corresponding expression values. If $X_i<X_j$ or $X_i>X_j$ for both patient A and B, the distance of these two patients was 0; otherwise the distance was 1. We summed up the distances for all 15 possible comparisons and recorded it as the total distance of patients A and B on network *P*. This procedure was repeated for all patients to get a 239 × 239 distance matrix.

Hierarchical clustering in MATLAB was used to **group the patients into two groups** (linkage method: weighted average distance, **Figure A.2**). The two largest branches were considered first; each group must have more than 10% of all samples (24 for this case) to be called a subtype, otherwise these samples are considered as outliers and the next largest group is taken as a possible subtype.

After two groups with reasonable sizes were determined, the Kaplan-Meier method was used to estimate the survival distributions. Log-rank tests were used to test the difference between survival groups.

REFERENCES

1. Holland, E.C., et al., *Astrocytes give rise to oligodendrogliomas and astrocytomas after gene transfer of polyoma virus middle T antigen in vivo.* American Journal of Pathology, 2000. **157**(3): p. 1031-1037.
2. Louis, D.N., et al., *The 2007 WHO classification of tumours of the central nervous system.* Acta Neuropathol, 2007. **114**(2): p. 97-109.
3. Nutt, C.L., et al., *Gene expression-based classification of malignant gliomas correlates better with survival than histological classification.* Cancer Res, 2003. **63**(7): p. 1602-7.
4. Shirahata, M., et al., *Gene expression-based molecular diagnostic system for malignant gliomas is superior to histological diagnosis.* Clinical Cancer Research, 2007. **13**(24): p. 7341-7356.
5. Kros, J.M., *Grading of Gliomas: The Road From Eminence to Evidence.* Journal of Neuropathology and Experimental Neurology, 2011. **70**(2): p. 101-109.
6. van den Bent, M.J., *Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective.* Acta Neuropathol, 2010. **120**(3): p. 297-304.
7. Bourne, T.D. and D. Schiff, *Update on molecular findings, management and outcome in low-grade gliomas.* Nature Reviews Neurology, 2010. **6**(12): p. 695-701.
8. Schneider, T., et al., *Gliomas in adults.* Dtsch Arztebl Int, 2010. **107**(45): p. 799-807; quiz 808.
9. Furnari, F.B., et al., *Malignant astrocytic glioma: genetics, biology, and paths to treatment.* Genes & Development, 2007. **21**(21): p. 2683-2710.
10. Freije, W.A., et al., *Gene expression profiling of gliomas strongly predicts survival.* Cancer Res, 2004. **64**(18): p. 6503-6510.
11. Phillips, H.S., 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**(3): p. 157-173.
12. Wong, K.K., et al., *Expression analysis of juvenile pilocytic astrocytomas by oligonucleotide microarray reveals two potential subgroups.* Cancer Res, 2005. **65**(1): p. 76-84.
13. Sun, L.X., et al., *Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain.* Cancer Cell, 2006. **9**(4): p. 287-300.
14. Eddy, J.A., et al., *Identifying Tightly Regulated and Variably Expressed Networks by Differential Rank Conservation (DIRAC).* Plos Computational Biology, 2010. **6**(5): p. -.
15. Bombonati, A. and D.C. Sgroi, *The molecular pathology of breast cancer progression.* Journal of Pathology, 2011. **223**(2): p. 307-317.
16. Krupp, W., et al., *Cytogenetic and molecular cytogenetic analyses in diffuse astrocytomas.* Cancer Genetics and Cytogenetics, 2004. **153**(1): p. 32-38.
17. Lee, Y., et al., *Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age.* Bmc Medical Genomics, 2008. **1**: p. -.
18. Kleihues, P. and H. Ohgaki, *Primary and secondary glioblastomas: from concept to clinical diagnosis.* Neuro Oncol, 1999. **1**(1): p. 44-51.
19. Ohgaki, H. and P. Kleihues, *Genetic pathways to primary and secondary glioblastoma.* American Journal of Pathology, 2007. **170**(5): p. 1445-53.
20. Maher, E.A., et al., *Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities.* Cancer Res, 2006. **66**(23): p. 11502-13.

21. Sirbu, A., H.J. Ruskin, and M. Crane, *Cross-platform microarray data normalisation for regulatory network inference.* PLoS One, 2010. **5**(11): p. e13822.
22. Wu, Z.J., et al., *A model-based background adjustment for oligonucleotide expression arrays.* Journal of the American Statistical Association, 2004. **99**(468): p. 909-917.
23. Subramanian, A., et al., *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15545-15550.
24. Cheng, Y., et al., *Pilocytic astrocytomas do not show most of the genetic changes commonly seen in diffuse astrocytomas.* Histopathology, 2000. **37**(5): p. 437-444.
25. Godard, S., 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**(20): p. 6613-6625.
26. Woo, H.G., et al., *Exploring Genomic Profiles of Hepatocellular Carcinoma.* Molecular Carcinogenesis, 2011. **50**(4): p. 235-243.
27. Grzmil, M. and B.A. Hemmings, *Deregulated signalling networks in human brain tumours.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2010. **1804**(3): p. 476-483.
28. Hermanson, M., et al., *Platelet-Derived Growth-Factor and Its Receptors in Human Glioma Tissue - Expression of Messenger-Rna and Protein Suggests the Presence of Autocrine and Paracrine Loops.* Cancer Res, 1992. **52**(11): p. 3213-3219.
29. Plate, K.H., et al., *Platelet-Derived Growth-Factor Receptor-Beta Is Induced during Tumor-Development and up-Regulated during Tumor Progression in Endothelial-Cells in Human Gliomas.* Laboratory Investigation, 1992. **67**(4): p. 529-534.
30. Calzolari, F. and P. Malatesta, *Recent Insights into PDGF-Induced Gliomagenesis.* Brain Pathology, 2010. **20**(3): p. 527-538.
31. Knobbe, C.B. and G. Reifenberger, *Genetic alterations and aberrant expression of genes related to the phosphatidyl-inositol-3 '-kinase/protein kinase B (Akt) signal transduction pathway in glioblastomas.* Brain Pathology, 2003. **13**(4): p. 507-518.
32. Onishi, M., et al., *Angiogenesis and invasion in glioma.* Brain Tumor Pathology, 2011. **28**(1): p. 13-24.
33. Gosselin, D. and S. Rivest, *Role of IL-1 and TNF in the brain: Twenty years of progress on a Dr. Jekyll/Mr. Hyde duality of the innate immune system.* Brain Behavior and Immunity, 2007. **21**(3): p. 281-289.
34. Vega, E.A., M.W. Graner, and J.H. Sampson, *Combating immunosuppression in glioma.* Future Oncology, 2008. **4**(3): p. 433-442.
35. Sciume, G., A. Santoni, and G. Bernardini, *Chemokines and glioma: Invasion and more.* Journal of Neuroimmunology, 2010. **224**(1-2): p. 8-12.
36. Hu, B., et al., *Angiopoietin 2 induces glioma cell invasion by stimulating matrix metalloprotease 2 expression through the alpha(v)beta(1) integrin and focal adhesion kinase signaling pathway.* Cancer Res, 2006. **66**(2): p. 775-783.
37. Strakova, N., et al., *Peroxisome proliferator-activated receptors (PPAR) agonists affect cell viability, apoptosis and expression of cell cycle related proteins in cell lines of glial brain tumors.* Neoplasma, 2005. **52**(2): p. 126-136.
38. Shannon, P., et al., *Pathological and molecular progression of astrocytomas in a GFAP:12 V-Ha-Ras mouse astrocytoma model.* American Journal of Pathology, 2005. **167**(3): p. 859-67.

39. Wu, P., et al., *Apollo contributes to G overhang maintenance and protects leading-end telomeres.* Mol Cell. **39**(4): p. 606-17.
40. Lenain, C., et al., *The Apollo 5' exonuclease functions together with TRF2 to protect telomeres from DNA repair.* Curr Biol, 2006. **16**(13): p. 1303-10.
41. Hejna, J., et al., *The hSNM1 protein is a DNA 5'-exonuclease.* Nucleic Acids Res, 2007. **35**(18): p. 6115-23.
42. Ye, J., et al., *TRF2 and apollo cooperate with topoisomerase 2alpha to protect human telomeres from replicative damage.* Cell, 2010. **142**(2): p. 230-42.
43. Pommier, Y., *Topoisomerase I inhibitors: camptothecins and beyond.* Nature Reviews Cancer, 2006. **6**(10): p. 789-802.
44. Capranico, G., J. Marinello, and L. Baranello, *Dissecting the transcriptional functions of human DNA topoisomerase I by selective inhibitors: Implications for physiological and therapeutic modulation of enzyme activity.* Biochimica Et Biophysica Acta-Reviews on Cancer, 2010. **1806**(2): p. 240-250.
45. Sordet, O., et al., *Topoisomerase I requirement for death receptor-induced apoptotic nuclear fission.* Journal of Biological Chemistry, 2008. **283**(34): p. 23200-23208.
46. Sasine, J.P., N. Savaraj, and L.G. Feun, *Topoisomerase I Inhibitors in the Treatment of Primary CNS Malignancies: An Update on Recent Trends.* Anti-Cancer Agents in Medicinal Chemistry, 2010. **10**(9): p. 683-696.
47. Kaidi, A., et al., *Human SIRT6 promotes DNA end resection through CtIP deacetylation.* Science, 2010. **329**(5997): p. 1348-53.
48. You, Z. and J.M. Bailis, *DNA damage and decisions: CtIP coordinates DNA repair and cell cycle checkpoints.* Trends Cell Biol, 2010. **20**(7): p. 402-9.
49. Schaeper, U., et al., *Interaction between a cellular protein that binds to the C-terminal region of adenovirus E1A (CtBP) and a novel cellular protein is disrupted by E1A through a conserved PLDLS motif.* J Biol Chem, 1998. **273**(15): p. 8549-52.
50. Nakamura, K., et al., *Collaborative action of Brca1 and CtIP in elimination of covalent modifications from double-strand breaks to facilitate subsequent break repair.* PLoS Genet. **6**(1): p. e1000828.
51. Galanty, Y., et al., *Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks.* Nature, 2009. **462**(7275): p. 935-9.
52. Fusco, C., A. Reymond, and A.S. Zervos, *Molecular cloning and characterization of a novel retinoblastoma-binding protein.* Genomics, 1998. **51**(3): p. 351-8.
53. Liu, F. and W.H. Lee, *CtIP activates its own and cyclin D1 promoters via the E2F/RB pathway during G1/S progression.* Mol Cell Biol, 2006. **26**(8): p. 3124-34.
54. Wu, M.H., et al., *CUP silencing as a novel mechanism of tamoxifen resistance in breast cancer.* Molecular Cancer Research, 2007. **5**(12): p. 1285-1295.
55. Menzel, T., et al., *A genetic screen identifies BRCA2 and PALB2 as key regulators of G2 checkpoint maintenance.* EMBO Rep, 2011. **12**(7): p. 705-12.
56. Wolffe, A.P., *Chromatin remodeling: why it is important in cancer.* Oncogene, 2001. **20**(24): p. 2988-90.
57. Davis, P.K. and R.K. Brachmann, *Chromatic remodeling and cancer.* Cancer Biology & Therapy, 2003. **2**(1): p. 22-29.
58. Ancelin, K., et al., *Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells.* Nat Cell Biol, 2006. **8**(6): p. 623-30.

59. Majumder, S., et al., *Methylation of histone H3 and H4 by PRMT5 regulates ribosomal RNA gene transcription.* J Cell Biochem. **109**(3): p. 553-63.
60. Jansson, M., et al., *Arginine methylation regulates the p53 response.* Nat Cell Biol, 2008. **10**(12): p. 1431-9.
61. Nozaki, M., et al., *Roles of the functional loss of p53 and other genes in astrocytoma tumorigenesis and progression.* Neuro Oncol, 1999. **1**(2): p. 124-37.
62. Meek, D.W., *Tumour suppression by p53: a role for the DNA damage response?* Nature Reviews Cancer, 2009. **9**(10): p. 714-23.
63. Senft, C., et al., *Inhibition of the JAK-2/STAT3 signaling pathway impedes the migratory and invasive potential of human glioblastoma cells.* J Neurooncol. **101**(3): p. 393-403.
64. Lo, H.W., et al., *Constitutively activated STAT3 frequently coexpresses with epidermal growth factor receptor in high-grade gliomas and targeting STAT3 sensitizes them to Iressa and alkylators.* Clin Cancer Res, 2008. **14**(19): p. 6042-54.
65. Halliday, G.M., et al., *SWI/SNF: a chromatin-remodelling complex with a role in carcinogenesis.* Int J Biochem Cell Biol, 2009. **41**(4): p. 725-8.
66. Reisman, D., S. Glaros, and E.A. Thompson, *The SWI/SNF complex and cancer.* Oncogene, 2009. **28**(14): p. 1653-68.
67. Ho, L., et al., *An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency.* Proc Natl Acad Sci U S A, 2009. **106**(13): p. 5181-6.
68. Ho, L., et al., *An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network.* Proc Natl Acad Sci U S A, 2009. **106**(13): p. 5187-91.
69. Yoo, A.S. and G.R. Crabtree, *ATP-dependent chromatin remodeling in neural development.* Curr Opin Neurobiol, 2009. **19**(2): p. 120-6.
70. Pal, S., et al., *Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes.* Molecular and Cellular Biology, 2004. **24**(21): p. 9630-45.
71. Zhao, Y., et al., *A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing.* Mol Cell, 2008. **29**(1): p. 92-101.
72. Basu, D., et al., *Development and validation of a fluorogenic assay to measure separase enzyme activity.* Analytical Biochemistry, 2009. **392**(2): p. 133-138.
73. Van Meir, E.G., et al., *Exciting New Advances in Neuro-Oncology The Avenue to a Cure for Malignant Glioma.* Ca-a Cancer Journal for Clinicians, 2010. **60**(3): p. 166-193.
74. Lee, Y., et al., *Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age.* BMC Med Genomics, 2008. **1**: p. 52.
75. Byrne, T.N., *Cognitive sequelae of brain tumor treatment.* Curr Opin Neurol, 2005. **18**(6): p. 662-6.
76. Campana, W.M. and R.R. Myers, *Erythropoietin and erythropoietin receptors in the peripheral nervous system: changes after nerve injury.* FASEB J, 2001. **15**(10): p. 1804-6.
77. Celik, M., et al., *Erythropoietin prevents motor neuron apoptosis and neurologic disability in experimental spinal cord ischemic injury.* Proc Natl Acad Sci U S A, 2002. **99**(4): p. 2258-63.
78. Kertesz, N., et al., *The role of erythropoietin in regulating angiogenesis.* Dev Biol, 2004. **276**(1): p. 101-10.

79. Keswani, S.C., et al., *A novel endogenous erythropoietin mediated pathway prevents axonal degeneration.* Ann Neurol, 2004. **56**(6): p. 815-26.
80. Konishi, Y., et al., *Trophic effect of erythropoietin and other hematopoietic factors on central cholinergic neurons in vitro and in vivo.* Brain Res, 1993. **609**(1-2): p. 29-35.
81. Shingo, T., et al., *Erythropoietin regulates the in vitro and in vivo production of neuronal progenitors by mammalian forebrain neural stem cells.* J Neurosci, 2001. **21**(24): p. 9733-43.
82. Erbayraktar, S., et al., *Carbamylated erythropoietin reduces radiosurgically-induced brain injury.* Mol Med, 2006. **12**(4-6): p. 74-80.
83. Pinel, S., et al., *Erythropoietin-induced reduction of hypoxia before and during fractionated irradiation contributes to improvement of radioresponse in human glioma xenografts.* Int J Radiat Oncol Biol Phys, 2004. **59**(1): p. 250-9.
84. Stuben, G., et al., *Recombinant human erythropoietin increases the radiosensitivity of xenografted human tumours in anaemic nude mice.* J Cancer Res Clin Oncol, 2001. **127**(6): p. 346-50.
85. Acs, G., et al., *Hypoxia-inducible erythropoietin signaling in squamous dysplasia and squamous cell carcinoma of the uterine cervix and its potential role in cervical carcinogenesis and tumor progression.* Am J Pathol, 2003. **162**(6): p. 1789-806.
86. Belenkov, A.I., et al., *Erythropoietin induces cancer cell resistance to ionizing radiation and to cisplatin.* Mol Cancer Ther, 2004. **3**(12): p. 1525-32.
87. Liu, W.M., et al., *Effect of haemopoietic growth factors on cancer cell lines and their role in chemosensitivity.* Oncogene, 2004. **23**(4): p. 981-90.
88. Kumar, S.M., et al., *Functional erythropoietin autocrine loop in melanoma.* Am J Pathol, 2005. **166**(3): p. 823-30.
89. Nico, B., et al., *Epo is involved in angiogenesis in human glioma.* J Neurooncol, 2011. **102**(1): p. 51-8.
90. Cao, Y., et al., *Erythropoietin Receptor Signaling Through STAT3 Is Required For Glioma Stem Cell Maintenance.* Genes Cancer, 2010. **1**(1): p. 50-61.
91. Hassouna, I., et al., *Erythropoietin augments survival of glioma cells after radiation and temozolomide.* Int J Radiat Oncol Biol Phys, 2008. **72**(3): p. 927-34.

APPENDIX

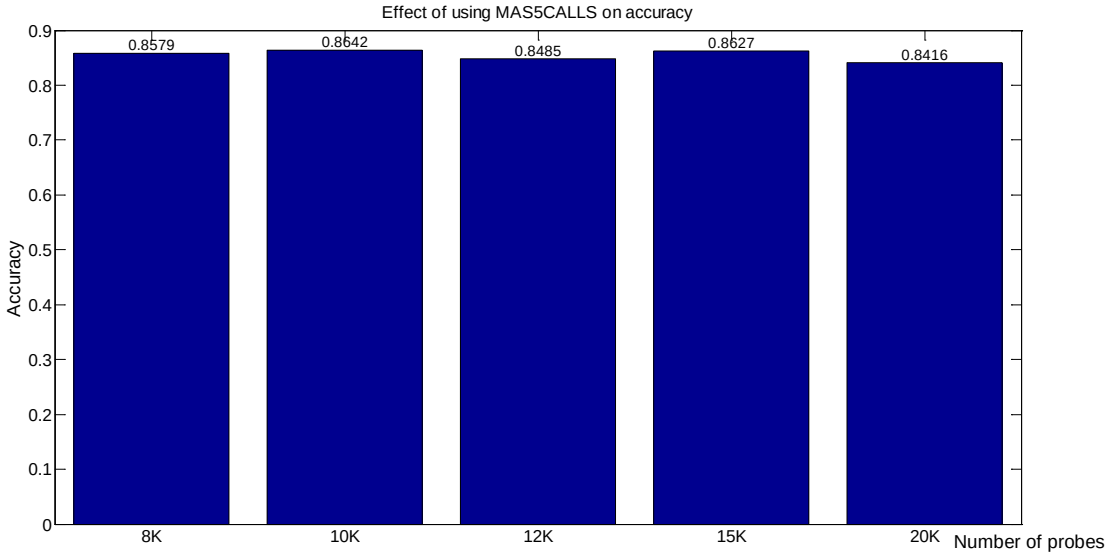


Figure A.1: Effect of filtering genes on accuracy. Average accuracies of 10 individual classifications by DIRAC do not vary significantly with number of genes present.

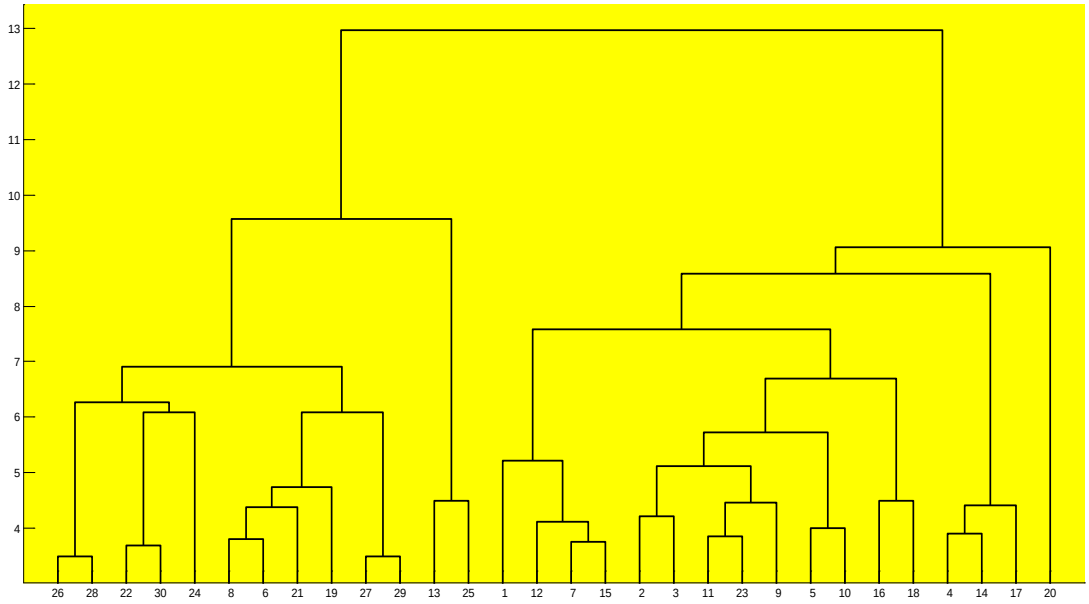


Figure A.2: Hierarchical clustering of HGA tumors. Two distinct clusters are formed; the distance matrix is defined by gene pair ordering matches in EPONFKB network.

Table A.1

Gene Name	Full Name/Atlas	Encoded protein function and association with cancer
AHCY	adenosylhomocysteinase	Hydrolase, upregulated in colorectal cancer
AK2	adenylate kinase 2	Induces apoptosis, upregulated in epilepsy patients
BCL2L11	BCL2-like 11/BIM	Induces apoptosis, upregulated in cancer cells
CANX	calnexin	Apoptosis and protein folding
DCLRE1B	DNA cross-link repair 1B/ APOLLO	DNA damage repair
ENY2	enhancer of yellow 2	Chromatin regulation
ESPL1	extra spindle pole bodies homolog 1	Chromatin regulation Oncogene, overexpressed in breast, prostate cancers and osteosarcoma.
GBE1	glucan (1,4-alpha-), branching enzyme 1	glycogen branching enzyme
GEMIN4	gem associated protein 4	part of a complex functioning in spliceosomal snRNP assembly in the cytoplasm
GNPAT	glyceronephosphate O-acyltransferase	essential to the synthesis of ether phospholipids; implicated in schizophrenia
GRM8	glutamate receptor, metabotropic 8	G protein-coupled receptors for excitatory neurotransmitter inhibition of the cyclic AMP cascade implicated in schizophrenia
KIF15	kinesin family member 15	neuronal development
NCAPG	non-SMC condensin I complex, subunit G	chromatin regulation
NUP107	nucleoporin 107kDa	Depletion induces apoptosis
NUP205	nucleoporin 205kDa	Essential component of nuclear pore complex
PALB2	partner and localizer of BRCA2	DNA damage repair
PDIA6	protein disulfide isomerase family A, member 6	folding of disulfide-bonded proteins, a biomarker for prostate cancer
PMS2L11	postmeiotic segregation increased 2 pseudogene 11	unknown function
PRMT5	protein arginine methyltransferase 5	chromatin remodeling
RBBP8	retinoblastoma binding protein 8/ CtIP	DNA damage repair
RSRC1	arginine/serine-rich coiled-coil 1	participate in multiple steps of mRNA splicing, implicated in schizophrenia
TFR2	transferrin receptor 2	involved in iron metabolism, hepatocyte function and erythrocyte differentiation
TMEM194A	transmembrane protein	unknown function

	194A	
TMEM45A	transmembrane protein 45A	unknown function
TOP1	topoisomerase (DNA) I	DNA damage repair
WDR76	WD repeat domain 76	Unknown function
ZNF282	zinc finger protein 282	Binds to a repressive element of the human T cell leukemia

Table A.1, name and functions of 27 monotonically increasing transcripts

Table A.2

Gene Name	Full Name/Atlas	Encoded protein function and association with cancer
AAK1	AP2 associated kinase 1	Regulation of clathrin-mediated endocytosis
ACSL4	acyl-CoA synthetase long-chain family member 4	Axonal transport Blocks apoptosis and promotes carcinogenesis
API5	apoptosis inhibitor 5	Inhibits apoptosis
C10orf97	chromosome-10, open reading frame-97	Pro-apoptosis, tumor-suppressor
C6orf211	chromosome 6 open reading frame 211	Unknown function
DSTYK	dual serine/threonine and tyrosine protein kinase	Induce apoptosis
GNAO1	guanine nucleotide binding protein (G protein), alpha activating activity polypeptide O	Regulation of cAMP levels Implicated in schizophrenia
ITM2B	integral membrane protein 2B	Induce apoptosis
NDUFB8	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 8,	Accessory subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase
NMT2	N-myristoyltransferase 2	Depletion induced apoptosis
PRNP	prion protein	neuronal development and synaptic plasticity, implicated in neurodegenerative diseases
RAB33B	RAB33B, member RAS oncogene family	Protein transport
RNF11	ring finger protein 11	May play a role in inflammatory pathways
SACM1L	SAC1 suppressor of actin mutations 1-like	Hydrolase
SDCCAG1	Serologically defined colon cancer antigen 1	Plays a role in nuclear export
SMARCA2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin,	Chromatin regulation DNA damage repair

	subfamily a, member 2/BRM	
STRN3	striatin, calmodulin binding protein 3	scaffolding or signaling
THOC7	THO complex 7 homolog	mRNA export
USP12	ubiquitin specific peptidase 12	Deubiquitinating enzyme, associated with Parkinson's disease

Table A.2, name and functions of 19 monotonically decreasing transcript

Network Name	P-value
EPONFKB	$2.5*10^{-8}$
CARDIACEGF	$2.7*10^{-5}$
IL22BP	$4.5*10^{-5}$
EPO	$4.6*10^{-5}$
FIBRINOLYSIS	$5.3*10^{-5}$

Table A.3: Networks that significantly differentiate 239 HGA patients into two groups with survival difference. These networks are arranged in decreasing p-values; the most significant EPONFKB outperform histological grade; the others did not but still significant)