Role of Long Noncoding RNAs in Aggressive Brain Tumors

Brian Joseph Reon

BS Biochemistry, Louisiana State University, 2008

MS Biochemistry, Louisiana State University, 2010

A dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for

the degree of Doctor of Philosophy

Experimental Pathology
University of Virginia
January, 2017

Table of Contents

Abstract

Long noncoding RNAs (lncRNAs) are emerging as key regulators of cellular processes and are commonly found to be misregulated in many human pathologies, including cancer. Glioblastomas (GBM) and lower-grade gliomas (Grade II and III) are aggressive, difficult to treat brain tumors that have high mortality rates. Although much of the research into glioma biology has focused on identifying the molecular drivers of oncogenesis, these efforts have primarily focused on alterations in protein coding genes and the role of lncRNAs has not been sufficiently characterized. This dissertation focuses on expanding our knowledge of lncRNAs in gliomas by analyzing their expression globally and more targeted studies examining the role of two lncRNAs.

In our initial study, we analyzed the expression data from over 750 RNA-seq datasets from GBMs, Grade III and II gliomas and normal brain tissue. We found that hundreds of lncRNAs are differentially expressed in gliomas compared to normal brain tissue. Furthermore, many lncRNAs were found to be preferentially expressed within certain GBM and LGG subtypes. Using these subtype specific lncRNAs we identified similarities between the highly aggressive IDH1/2 wt LGG subtype and mesenchymal GBMs. We also used Cox regression to create a survival algorithm that is capable of separating LGG patients into two distinct prognostic groups. Lastly, we identified all lncRNAs that are associated with GBM patient survival, to aid in identifying which lncRNAs might play critical roles in brain tumors.

Using our brain tumor lncRNA expression and survival association data, we identified two oncogenic lncRNAs, GS1-124K5.4 and LINC00152, for further study. Both LINC00152 and GS1-124K5.4 are upregulated in GBMs and aggressive gliomas, and high expression of either lncRNAs is associated with negative patient outcomes in GBMs. High levels of LINC00152 has

no affect on cell growth but lead to an increase invasion in U87 cells. In contrast, expression of GS1-124K5.4 does not affect invasion but high expression of GS1-124K5.4 leads to increased cell growth. Secondary structure analysis of LINC00152 suggests that a protein-bound stem-loop in the 3' end of LINC00152 is partially involved in LINC00152's proinvasive function. Furthermore, LINC00152 is upregulated in 10 other tumors types and high expression of is associated with a poor prognosis in 7 other tumors. These results suggest that LINC00152 potentially functions as an oncogene in many cancers.

Acknowledgements

I would like to start off by thanking my parents, AJ and Cathy Reon. Without your nurturing and patience, none of this would have been possible. Both of you have been amazing role models for me and Erin and I try my best to follow your example of what it means to be a spouse, parent and friend. Although I am sure I have shaved some years off of each of your lives, I am forever grateful for all that y'all have taught me.

I would also like to thank my thesis committee for help guiding me with my projects and your feedback throughout my thesis research.

I also want to thank my mentor, Anindya, for all you have done for me. I have enjoyed our many conversations, which have ranged from science to politics. Your passion for science is unmistakable and infectious. Although I probably won't miss Monday lab meetings, I will miss our frequent conversations. I have learned so much under your tutelage about science and what it takes to run a successful lab and I hope to take some of those lessons with me for the next stage in my career.

Last and most importantly, I would like to thank my wife, Jessica, and our three children, Owen, Julianne and Elizabeth for their eternal patience and support for me during my academic training. Each of you has helped me so much during this process and have been huge positive motivating forces in my life. Y'all help remind me what is truly important in life. No matter how tough a day I might have had, coming home and being greeted with three running hugs always makes it better. Thanks for everything!

Chapter 1: Introduction

Gliomas

Although brain and CNS (Central Nervous System) tumors have the 7th highest incidence rate of all adult tumors, brain and CNS tumors are the most common cancer in children and adolescents, with the second highest mortality rate in this age group, second only to leukemia [1]. Brain and CNS tumors, as well as other tumor types, can broadly be separated into two groups based on their malignant potential. Non-malignant brain and CNS tumors are benign tumors that, as described by the CBTRUS (Central Brain Tumor Registry of the United States), have an ICDO (International Coding for Diseases of Oncology) morphology behavior score of "0" or "1", representing benign or uncertain behavior, respectively [1]. Of the malignant primary brain tumors that occur in adults, malignant gliomas are the most common, making up over ¾ of all newly diagnosed cases in the United States [1].

Gliomas are a group of tumors, including glioblastoma, astrocytoma, ependymoma, oligodendroglioma, mixed glioma as well as other rarer tumor types, that are believed to be derived from glial cells. Glial cells are the non-neuronal supporting cells of the central and peripheral nervous system. In the CNS, they play several critical roles in maintaining neurologic function and include astrocytes, microglia and oligodendrocytes. Oligodendrocytes are the electrical insulators of the CNS, forming lipid rich myelin sheaths around neuronal axons, in addition to releasing other trophic factor to support neuronal function [2]. Astrocytes, named after their star like appearance, aid in the maintenance of the blood-brain barrier, in the development of neurologic synapses as well as maintenance and elimination of existing synapses [3]. Microglia are the resident macrophages of the CNS and serve as the main immune cell in the brain, as the brain is traditionally considered to be "immune privileged" and relatively devoid of other common immune cells [4].

Glioma Cell of Origin

As mentioned above, the glioma class of brain and CNS tumors consists of multiple glia derived tumors, whose names, astrocytoma, oligodendroglioma, etc, originate from the resemblance of each of the tumor types to the corresponding glial cell population, i.e. astrocytes and oligodendrocytes, respectively. This association goes beyond mere histological appearances, as these tumors share many of the same cellular markers as the glial cell populations, such as GFAP (glial fibrillary acidic protein), OLIG2 (oligodendrocyte transcription factor 2) and NG2 (neural/glial antigen 2) [5–7]. While these correlations have fueled much work speculating the cell of origin for the respective tumors, these connections are not as easily drawn for mixed tumor types such as oligoastrocytoma, which have both astrocytic and oligodendocytic morphologic components [8]. Recent work has focused on two stem cell populations of the brain, NSCs (neural stem cells) and OPCs (oligodendrocyte progenitor cells) as potential cells of origin for gliomas. Mouse models that specifically inactivate tumor suppressors, including TP53 and PTEN, in distinct cell populations or brain region, have provided evidence that both NSC and OPCs can serve as the cell of origin for many gliomas [9–11].

Symptoms and Risk Factors of Gliomas

Due to difficulties of performing routine surveillance on the brain, malignant gliomas are not typically detected until later in the disease process and commonly present with nonspecific symptoms, such as nausea, headaches and changes in cognitive functions [12]. The most common presenting symptom in GBMs is headache, while seizures are the most common symptom of grade III gliomas. However, depending on the location within the brain, malignant gliomas can present with more specific, localized symptoms, like changes in vision, gait or speech patterns [12]. It should be noted that most malignant gliomas are located supratentorially, with 40.5% of all

malignant brain and CNS tumors arising in the frontal and temporal lobes [1]. In rare cases, patients can present with back pain or cranial nerve palsies if the tumor has disseminated throughout the meninges [13].

Many studies have attempted to identify environmental risk factors that are associated with the development of GBMs. One potential risk factor that received a great deal of publicity in the lay press, is cell phone usage. Although a few of the initial studies examining the relationship between cell phone use and gliomas suggested that there was in fact an association, this correlation was later attributed to bias [14,15]. More recent work, including the largest study to date, has found that there is no significant increased risk in developing gliomas with cell phone usage [16]. There have also been suggestions that gliomas are associated with N-nitroso containing food products, head injuries and occupational exposures, but the evidence supporting these claims are ambiguous [17]. Exposure to ionizing radiation (IR) is the only risk factor that has been proven to be associated with glioma development [18]. The most tragic and famous example of this is the higher incidence of gliomas and meningiomas in survivors of the nuclear bomb explosion in Hiroshima, which correlated to the level of exposure to IR [19,20]. There is also an increased risk of developing a glioma with exposure to therapeutic doses of IR [20]. Interestingly, multiple studies have found an inverse correlation between having a history of allergies and high titers of serum IgE with the development of gliomas [21–24].

In addition to environmental risk factors, a few hereditary syndromes are associated with a higher risk of developing gliomas, such as Turcot Syndrome and type 1 and type 2 neurofibromatosis [18,25]. These conditions are however rarely seen in glioma patients, and some studies report less than 1% of glioma patients were diagnosed with a hereditary syndrome [26]. Furthermore, glioma patients do not typically have a family history of gliomas. GWAS (genome

wide association studies) have found links of glioma formation to SNPs (single nucleotide polymorphisms) in genes involved in DNA repair pathways and cell cycle regulation, including ERCC1, ERCC2, MGMT and PRKDC [27–30]. However these studies were carried out on relatively small sample sizes and future work is needed to determine the reliability of the GWAS gene predictions.

Lower-Grade Gliomas

The Cancer Genome Atlas (TCGA) designated grade II and III gliomas as lower-grade gliomas (LGG) [31]. It should be pointed out that a grade III glioma is by no mean "low grade", however for the purposes of this dissertation, it will be grouped with grade II gliomas and called LGGs. LGGs are a composite of multiple histological classes, including the ones mentioned earlier (astrocytoma, oligodendroglioma etc.), and prognosis within the same grade of gliomas can vary greatly [1]. IDH1 and IDH2 (IDH1/2) mutations are the most common genomic alteration found in LGGs, with nearly 80 % of LGGs having a mutation in either gene [31]. Mutations to IDH1/2 are exclusively found in the active site of isocitrate dehydrogenase and lead to a gain of function and the production of the oncometabolite 2-hydroxyglutarate (2HG) instead of the normal Krebs cycle metabolite, α-ketoglutarate (αKG) [32]. 2HG acts as an oncometabolite by inhibiting enzymes within the cell that normally use αKG as a substrate, and this inhibition leads to oncogenic transformation of mutated cells [32].

Beyond tumor grade, scientists have determined that LGGs can be separated into three distinct subgroups based on 2 commonly occurring genomic alteration in gliomas, IDH1/2 mutational status and codeletion of 1p19q [33]. These subgroups, IDH 1/2 wild type (IDH 1/2 wt), IDH 1/2 mutant (IDH 1/2 mut), and IDH 1/2 mut 1p19q codeletion (IDH 1/2 mut 1p19q codel), were also independently identified by integrating multiple high throughput sequencing/array

modalities which assessed mRNA and microRNA levels, DNA methylation and DNA copy number [31]. Through these sequencing efforts, we now know that each of the LGG subtypes are associated with different histological classes and groups of mutations. For instance, the IDH 1/2 mut 1p19q codel are most frequently found in oligodendrogliomas and disproportionately have mutations in CIC and FUBP1 as well as activation of telomerase [31]. The IDH 1/2 mut subtype is more commonly found in astrocytomas and oligoastrocytomas and almost invariably have deactivating mutations in the p53 pathway and are rarely found to have TERT activation. This is in contrast to IDH 1/2 mut 1p19q codel gliomas, suggesting a possible mutational divergence from an initial IDH 1/2 mutated cell. Also, IDH 1/2 wt gliomas have a higher mutational burden compared to the other LGG subtypes, and more closely resemble the mutational spectrum seen in GBMs, with activation of the EGFR pathway and deactivation of PTEN. Importantly, the subtypes are associated with different patient outcomes, with IDH 1/2 mut 1p19q codel gliomas having the best prognosis, median survival of 8 years, and IDH 1/2 wt having the worst prognosis, median survival, 1.7 years [31].

Glioblastoma

Glioblastoma is the most common malignant brain tumor, comprising 46% of all malignant brain and CNS cancers [1]. The incidence rate of glioblastoma is highest in Caucasians, with an incidence rate of 3.45/100,000, over two fold higher than the nearest race [34]. GBM are highly invasive tumors and invariably recur following standard treatment. GBMs have a very poor prognosis, with a median survival time of 1 year [35]. Unlike other tumor types such as colon cancer or breast cancer, whose increasing disease stages are viewed as being part of a disease continuum, *i.e.* stage 1 tumors progress to stage 2 tumors which then progresses to stage 3 and so

on, GBMs are not typically viewed in this way [31]. As mention earlier, IDH1/2 are mutated in 80% of all LGGs whereas they are rarely found to be mutated in primary GBMs [31].

While tumors of a given cancer type were once thought of as a homogeneous disease, we now know that there is considerable heterogeneity between tumors as well as within tumors. Indeed, through high throughput sequencing efforts led by the TCGA, it was discovered that in GBMs, there are 4 distinct tumor subtypes, termed Neural, Proneural, Classical and Mesenchymal. In addition to differences in transcriptional profiles, each GBM subtype is associated with specific mutations. Over half of proneural GBMs have mutations of p53 and are more likely to have mutations in IDH1/2. Interestingly, classical GBMs are devoid of p53 mutations, but often have the activating EGFR rearrangement EGFRvIII. Mesenchymal GBMs are more likely to have mutations in NF1, similar to IDHwt LGGs [31,36]. Importantly, these subtypes are associated with clinical parameters, such as response to intensive therapy. Classical and mesenchymal GBMs have prolonged survival following intensive therapy, where as proneural and neural GBMs do not have better response rates following aggressive therapy. More work is needed to understand if the GBM subtypes display preferential sensitivity to targeted therapies.

Treatment of GBMs

Following diagnosis, standard treatment for GBMs consists of maximal surgical resection, in cases where surgical resection is possible (due to location in the brain), adjuvant radiation therapy and treatment with Temozolomide, a DNA alkylating agent [37,38]. Response to temozolomide is strongly associated with methylation at the MGMt (O-6-methylguanine-DNA methyltransferase) promoter, a gene which encodes a protein involved in repairing the alkylated DNA adducts formed by temozolomide [39]. Patients with methylated MGMT promoters have much better outcomes with temozolomide treatment compared to unmethylated MGMT, with a

two-year survival of 49% and 15%, respectively [40]. The standard of care has extended overall survival for GBM patients, but patients almost always have tumor recurrence and the 5 year overall survival is 10% [40].

Other treatments have been tested in GBM patients, with mixed results. The use of a portable device that produces low-intensity alternating electric fields directed towards the original tumor location has been examined for patients with recurrent and primary GBMs, and although controversial, was shown to increase overall survival and progression-free survival [41]. Other studies have examined the use of carmustine wafers, a DNA interstrand crosslinking agent, at the surgical site, but the clinical trial have shown no statistically significant increase in overall survival for GBM patients [42]. In another string of clinical trials, bevacizumab, the anti-VEGF antibody, initially showed promise in a set of phase II clinical trials [43,44], but these results were not validated in subsequent randomized trials and as a result is not recommended for use in newly diagnosed GBM patients [45,46]. Although new therapeutics are in development and new clinical trials are ongoing, the targeted therapeutic approaches have not proved extremely effective. It is important to investigate emerging fields of biology to possibly identify novel therapeutic targets or to find tumor markers that single out tumors that are susceptible to current or novel therapies.

ncRNAs

The central dogma of molecular biology posited that there is a natural flow of genetic information, from DNA to RNA and ultimately leading to the production of proteins. Proteins were long thought to be the major functional effectors of the cell until relatively recently when the role of RNAs, other than rRNAs and tRNAs, have begun to be appreciated for their roles in various cellular processes. Exceptions to the central dogma began to appear in the 80's, when it was discovered that a large portion of the nuclear RNAs had 5' caps and were polyadenylated but did

not contribute to the cell's pool of coding transcripts (1). In the early eighties, the discovery of two non-coding transcripts, H19 and XIST, stirred up a great deal of excitement for the role of ncRNAs in normal cellular processes (2,3). However, over the next decade following their discovery, relatively little progress was made in the way of identifying and characterizing new non-coding RNAs. The advent of high-throughput sequencing transformed basic science research and in particular, truly revolutionized the noncoding RNA field. Through the ENCODE (Encyclopedia of DNA Elements) project as well as other global sequencing initiatives, it became clear that up to 76% of the mappable human genome is transcribed and a large number of these transcripts do not code for proteins (4,5,6). While the exact number of ncRNAs is debated, these sequencing efforts have shown that ncRNAs are ubiquitously expressed, with numbers rivaling those of coding RNAs. As the examples of ncRNAs increased, they began to be subdivided into distinct categories. One subcategory is long non-coding RNAs (lncRNAs), somewhat arbitrarily defined as RNA species greater than 200 bp, which do not code for proteins (7). LncRNAs can be further subdivided based on their relative position to neighboring genes. Antisense lncRNAs are lncRNAs that are transcribed from the opposite strand of a protein-coding gene, lncRNAs located within the introns of protein coding genes are termed intronic lncRNAs, and lncRNAs that lie in regions of the genome between genes are called lincRNAs (long intervening non-coding RNAs). More recently, new sub-classes of lncRNAs have emerged called eRNAs (enhancer RNAs) and circRNAs (circular RNAs).

lncRNAs

The newfound enthusiasm for the lncRNA field was not without controversy. Several skeptics pointed out that a large fraction of the ncRNA transcripts identified in high-throughput datasets were not evolutionarily conserved and could be artifacts of transcriptional noise that serve

no functional role in cells (9). As criticism mounted, a landmark paper by Guttman *et al.* sought to tackle this issue by applying similar standards used for protein coding genes to lncRNAs. Coding genes were known to have a "H3K4-K36 signature" pattern, where H3K4me3 marks gene promoters and H3K36me3 covers gene bodies (10). When these same genomic features were used to define intergenic lincRNAs, Guttman *et al.* found that these lncRNAs were in fact evolutionarily conserved to a greater extent than 5' and 3' UTRs of protein coding genes, although less so than the coding regions of mRNA (11). However, the relatively lower level sequence conservation of lncRNAs is most likely due to the fact that RNA secondary structure is more resilient to mutations than that of protein coding genes. This allows lncRNAs a greater degree of evolutionary flexibility and could more rapidly give rise to new functional roles in the cell.

Although lncRNAs do not produce functional proteins, it is quite common for a lncRNA to have an open reading frame (ORF). LncRNAs are on average 500 bp in length and a large percentage of lncRNAs are over 2 kb. By random chance, a 2 kb lncRNA will have, on average, a 200nt ORF. It is therefore critical to distinguish between true protein coding ORFs and those ORFs that are not translated and are only present by chance. There are several characteristics of protein coding genes that help distinguish between these two possibilities. Firstly, protein-coding genes will typically have annotated protein domains and the translated ORFs will usually share sequence homology to proteins in protein sequence databases. The second metric used is based on the observation that in protein coding regions, the evolutionary mutations in codons are not randomly distributed (i.e. there is a bias for silent substitutions), whereas the position of mutations in lncRNAs are randomly distributed within a given ORF (12). Over the years, several computational tools and sequencing assays have been developed to help determine the coding potential of a lncRNA [47,48]. One tool that has recently been adopted by the lncRNA field to measure the

coding potential of a lncRNA *in silico* is CPAT (coding potential assessment tool). CPAT combines 4 metrics: 1. ORF length, 2. ORF length relative to size of transcript, 3. Codon bias, 4. Hexamer usage bias (i.e. how frequently one sees two adjacent codons in a protein coding transcript). This tool has proven extremely effective in predicting the coding potential of a lncRNA [48].

High throughput sequencing of polysome associated RNAs found that many lncRNAs actually interact with the ribosome and this led many to speculate about the coding potential of lncRNAs. Newer methodologies used ribosome RNase footprinting followed by high throughput sequencing, Ribo-seq (Ribosome profiling), to find not only which RNAs are interacting with ribosome but how ribosomes are located along the transcript. This is important as actively translated mRNA have ribosomes preferentially located over the coding region and sequencing read peaks are staggered at 3 nt (nucleotide) intervals, consistent with the size of a codon. Analysis of Ribo-seq data has shown that although some lncRNAs do in fact interact with ribosomes, the pattern of ribosome association is not consistent with that of actively translated RNAs.

As mentioned earlier, the two classic examples of lncRNAs are Xist and H19. H19 is a 2.3 kb lncRNA which is highly expressed in the developing embryo in endodermal and mesodermal lineages. After birth, H19 expression is rapidly shut down and residual H19 transcripts are only detected in skeletal muscle. Interestingly, the paternal H19 locus is imprinted during development and H19 is therefore only transcribed from the maternal allele (2). Recently, studies have shown that H19 also encodes a microRNA, miR-675, which is involved in muscle development and differentiation (13). Perhaps the most studied lncRNA to date is Xist, the "Mother of all lncRNAs". Xist is a 17- to 20-kb lncRNA involved in X-chromosome inactivation, which allows for equal expression of X-linked relative to autosomal genes in both males and females, although the latter

have two copies of the X chromosome per cell. The process of X-chromosome inactivation is a beautiful example of the interplay between multiple lncRNAs that act in both cis- and trans-. Early during development, a short 1.6 kb Xist isoform, RepA, is produced and recruits PRC2, a repressive histone methyltransferase complex, to the Xist promoter (14). Interestingly, PRC2 is unable to bind to the Xist promoter due to the production of Tsix, Xist's antisense transcript. It is only when Tsix expression is repressed, that PRC2 is able to dock to the Xist promoter. PRC2 docking is also dependent on the presence of the transcription factor YY1, which is only localized on the future inactive X-chromosome (15). The trans- acting lncRNA Jpx is then recruited to the Xist promoter and promotes the expression of Xist (16). Newly transcribed Xist then binds PRC2 and deposits repressive methylation marks along the length of the inactive X-chromosome, shutting down gene expression (14). Other work has suggested that Xist interacts with SHARP and silences transcription through HDAC3 [49].

Due to the explosion in next generation sequencing, the number of predicted lncRNAs quickly increased, while the reports of characterized lncRNAs lagged behind. This began to change in the 2010's, when multiple studies began to appear in the literature that showed detailed mechanisms as well as associations between lncRNAs and a wide variety of pathologies. Through these functional characterizations, the functional diversity of lncRNAs has begun to be fully appreciated. It is helpful to think of the various ways a lncRNA might function based on the cellular compartment that it preferentially resides in. Broadly speaking, one can separate lncRNAs into mainly nuclear or mainly cytoplasmic, and I will discuss mechanistic examples of each in the following sections.

Function of Nuclear LncRNAs

Nuclear lncRNAs are traditionally thought of as gene expression regulators that carry out

their functions in cis or in trans through recruiting chromatin modifying complexes to target gene promoters. One of the most iconic lncRNA is HOTAIR, which is best known for being the first and best characterized example of an lncRNA playing a role in oncogenesis. HOTAIR is a lincRNA transcribed from the HOX loci that was first identified as being drastically deregulated in breast tumors [50]. HOTAIR is significantly overexpressed in breast cancer metastases and its expression level is associated with overall survival and risk of secondary metastatic events. *In vitro* studies have shown that HOTAIR interacts with PRC2 and promotes invasion in breast cancer cell lines. When overexpressed in breast cancer cells, HOTAIR is capable of relocating PRC2 on a genome-wide scale and induces a transcriptional profile similar to a more undifferentiated phenotype (17). Many other reports have shown that in addition to PRC2, lncRNAs interact with a wide variety of chromatin modifying complexes, such as PRC1, AR (androgen receptor) and LSD1, to regulate target gene expression [51](18,19). Initially, it was argued that nuclear lncRNAs primarily act in cis, only affecting the expression of genes in relative close proximity to their transcription site. This assertion was based on observations that there is significant correlation between the expression levels of lncRNAs and their adjacent protein-coding genes (20). However, subsequent studies have found that the correlations in gene expression seen with lincRNAs are similar to those found with protein-coding genes (21). In fact, a recent report which examined the expression profiles of embryonic stem cells following knockdown of over 140 lincRNAs showed that the vast majority of the lincRNAs (~95%) had minimal effects on neighboring gene expression, suggesting that most lncRNAs function in trans (22).

Although many of the characterized primarily nuclear lncRNAs function in this way, there is a growing body of evidence that lncRNAs function in other capacities, including acting as a molecular decoy for steroid receptors, regulating gene expression by promoting chromatin looping,

and altering mRNA splicing [52–54]. GAS5 (growth arrest-specific 5 noncoding RNA) is highly upregulated in serum- or nutrient-starved growth arrested cells [55]. Interestingly, GAS5 up-regulation is important to increase the cells susceptibility to undergo apoptosis. Through a yeast two-hybrid screen, the 5' end of GAS5 was found to interact with the DBD (DNA binding domain) of GR (Glucocorticoid receptor), which is also a key player in apoptotic cell fate decisions [52,56]. Further supporting this association, cytoplasmic GAS5 translocates to nucleus along with GR following treatment with dexamethasone. Surprisingly, the secondary structure of GAS5 is predicted to form several stem loops that resemble a GRE (glucocorticoid response element), which is critical for the GAS5-GR interaction. Through this interaction, GAS5 acts as a GRE decoy and inhibits many GR regulated genes [52]. Another function of lncRNAs is facilitating and augmenting chromosomal architecture. FIRRE (functional intergenic repeating RNA element) was first identified in a loss-of-function screen for lncRNAs involved in adipogenesis, and it is encoded on the X-chromosome, where it escapes X-chromosome inactivation [57]. FIRRE is localized in the nucleus and forms a roughly 5 MB domain surrounding its own gene locus. FIRRE also interacts with 5 other gene loci on different chromosomes, however these chromosomal loci are in close 3-dimensional proximity in the nucleus. Importantly, the spatial proximity of these disparate gene loci is dependent on FIRRE and hnRNPU, a nuclear matrix protein that interacts with FIRRE in one of its 156 base repeating RNA domains [54]. A separate example of a nuclear lncRNA function is asFGFR2, an antisense lncRNA that is on the opposite strand of FGFR2 (Fibroblast growth factor receptor 2). FGFR2 displays cell type-specific splicing patterns, with exonIIIb included in epithelial cells but excluded in mesenchymal cells. The inclusion of exonIIIb in epithelial cells is dependent on asFGFR2, which when expressed, interacts with and recruits PRC2

and KDM2a to the FGFR2 loci and deposit H3K27me3 along the gene body, leading to a chromatin environment that facilitates the inclusion of exonIIIb [53].

Functions of Cytoplasmic LncRNAs

There are many examples of lncRNAs that function in diverse cellular processes in the cytoplasm. As with mainly nuclear lncRNAs most of the characterized cytoplasmic lncRNAs carry out their function through base-pairing with a target RNA, although several notable exceptions exist (REF). Perhaps the best known role for cytoplasmic lncRNAs is acting as ceRNAs (competitive endogenous RNAs), by titrating away microRNAs from other transcriptional targets [58]. More recently, newer functions of cytoplasmic lncRNAs have been characterized, including increasing the efficiency of target mRNA translation, preventing ribosome binding on target mRNAs, increasing mRNA half-life and inhibiting signaling pathways through steric hindrance of signaling substrates [59–61].

Uchl1-as is an antisense lncRNA that partly overlaps with the protein coding gene Uchl1, that is highly expressed in ventral midbrain of mice and is downregulated in both *in vitro* and *in vivo* models of Parkinson's disease [61,62]. Uchl1-as is primarily located in the nucleus under basal conditions, however under stress conditions, Uchl1-as translocates to the cytoplasm and base pairs with a complementary region in the 5'end of Uchl1. This interaction enhances the translational efficiency of Uch1 mRNA by promoting polysome formation through an embedded inverted SINEB2 element on Uchl1-as [61]. Another example of a lncRNA effecting protein levels of a target gene is lincRNA-p21, which is a 3 kb lncRNA located 15 kb upstream of p21 [63]. lincRNA-p21 interacts with target RNAs including CTNNB1 and JUNB and decreases their protein levels by lowering their association with polysomes in a Rck dependent manner [64]. lincRNA-p21 levels are carefully regulated by the RNA binding protein HuR, which recruits

Ago2-let7, leading to the lincRNA-p21degradation [64].

One unique example of a cytoplasmic lncRNA that does not function through base-pairing with a target RNA is NKILA (NF-kB interacting lncRNA) [59]. NKILA is upregulated in response to many inflammatory stimuli that act through the NF-kB pathway. Interestingly, through binding p65, NKILA interacts with the NF-kB-IkB signaling complex. By binding this signaling complex, NKILA blocks the IKK phosphorylation site of IkB and inhibits the phosphorylation. In this way, NKILA plays an important role in a negative feedback loop to tamp down NF-kB signaling, following an initial boost in signaling. Furthermore, NKILA is commonly repressed in breast IDC (invasive ductal carcinoma) and patients with lower expression of NKILA have a worse prognosis compared to patients with high expression of NKILA [59].

TINCR (terminal differentiation-induced ncRNA) is a lncRNA that is critical in the epidermal differentiation of human keratinocytes, that was originally identified through a comparative transcriptomic screen in differentiating keratinocytes [60]. TINCR is primarily located in the cytoplasm (over 80% present in the cytoplasm) and knockdown of TINCR causes a dramatic decrease in keratinocyte differentiation. TINCR's role in epidermal differentiation is mediated by STAU1 (staufen1), an RNA binding protein involved in mRNA turnover (REF). Global analysis of RNAs that interact with TINCR found a 25-nt "TINCR-box" motif that was enriched in all TINCR-STAU1 interacting RNAs. The TINCR-STAU1 complex promoted increased stability of TINCR-box containing target RNAs, including some mRNAs that are known to function in epidermal differentiation [60]. TINCR further illustrates the functional versatility of lncRNAs as a novel medium for cellular regulation. In addition to the canonical lncRNAs mentioned above, newer classes of lncRNAs have been identified that play a role in many biological properties, including circRNAs and eRNAs.

eRNAs

eRNAs are noncoding RNAs that are transcribed from enhancers, defined as regions of the genome distal to coding genes that are occupied by p300 and RNA polymerase II, and typically contain stretches of H3K4 mono-methylation (H3K4me1). Recent transcriptome studies have found that many of these regions, which were long thought to be transcriptionally silent, are actually sites of active transcription to produce eRNAs. In contrast to typical lncRNAs, eRNAs mostly lack a poly-A tail, although this is not a steadfast rule. Furthermore, many eRNAs display changes in expression in response to specific stimuli, suggesting that they may not be transcriptional noise and in fact have regulatory capacity [65]. eRNAs have been shown to function in a couple of contexts, although it should be mentioned that it can be experimentally difficult to separate the function of the actual enhancer and the eRNAs that are transcribed from them. Examining '*de novo*' enhancers, Kaikkonen *et al.* demonstrated that the act of an RNA polymerase traversing an enhancer is necessary to deposit the requisite "enhancer" chromatin marks, independent of an actual transcript being produced initially [66]. Other work has shown a direct role of the eRNA transcript itself in regulating its neighboring gene expression by promoting the formation of chromatin loops through interactions with cohesion and the mediator complex [67,68].

MUNC (MyoD upstream noncoding RNA) is a ncRNA that arises from the DRR (distal regulatory region) of the key muscle transcription factor MyoD, and falls under the conventional definition of an eRNA [69,70]. Interestingly, knockdown of MUNC leads to a decrease in MyoD RNA. However MyodD RNA could be induced when MUNC was exogenously overexpressed, arguing that MUNC can function *in trans*. Furthermore, knockdown or overexpression of MUNC causes changes in a large array of genes, which further argues for MUNC acting in trans [69]. The

16

trans action of eRNAs is not purely limited to MUNC, as another group identified KLK3-eRNA, an eRNA from KLK3 (kallikrein related peptidase 3) enhancer that functions in trans [71]. MUNC and KLK3-eRNA, illustrates the issue of the definition of eRNAs being purely geographic (i.e. arise from genomic enhancers) and not necessarily being coupled to its function. In this way, the functional distinctions between some "eRNAs" and "lncRNAs" can become difficult to ascertain [69,70]. Indeed, there are new attempts to subdivide eRNAs into three categories based on their function; class I eRNAs have no known function associated with the act of their transcription or the RNA molecule derived from the enhancer; class II eRNAs are eRNAs whose function is solely derived from their transcription; class III eRNAs functions are mediated through the actual eRNA transcript [72]. More work is needed to fully understand the functional distinctions between "lncRNAs" and "eRNAs".

circRNAs

CircRNAs (circular RNAs) are a species of covalently closed lncRNA that originates from RNA polymerase II pre-mRNAs transcript back-splicing events [73]. A few examples of circRNAs were first described in the early 90s and thought to be rare byproducts of faulty splicing events [73–75]. More recently, high throughput sequencing of non-polyadenylated RNAs has found that circRNAs are more abundant then originally believed [76–78]. Some have argued that because circRNAs are present at such low levels, they don't have a functional role in the cell and are merely erroneous back-splicing of exons that occur stochastically during pre-mRNA processing. Several lines of evidence argue against these claims. Firstly, although it is true that most circRNAs are present at low levels, some circRNAs are highly expressed and are actually present at a higher abundance than their "properly" processed mRNA counterparts [79]. Secondly, the abundance of a given circRNA is not correlated with the levels of its parental transcript, arguing against

circRNAs being produced through stochastic processes [79–81]. Lastly, many circRNAs display cell-type-specific expression patterns that are comparable to the levels of cell-type specific expression seen in mRNAs [79,82].

Even though circRNAs are derived from mRNA exons and present primarily in the cytoplasm, endogenous circRNAs have not been found to be translated or associated with the ribosome [81]. The functional relevance of circRNAs has long been questioned, but recent work has identified a role for circRNAs in sponging microRNAs away from their parental transcripts, and presumably other transcripts with seed sequences to the sponged microRNA [76]. ciRS-7 is a circular RNA derived from CDR1 (cerebellar degeneration protein 1), and contains more than 60 binding sites for mir-7 and is strongly associated with Ago. ciRS-7 is highly expressed in the brain of mice and overexpression ciRS-7 in the brain of zebrafish leads to defects in midbrain development [76]. Other examples of circRNAs have shown that they can promote transcription in *cis* and in *trans* [83]. circRNAs represent a new field of lncRNA biology with promising implications for normal cell physiology and pathology, although more work is needed to determine how pervasive the functional relevance of circRNAs actually is.

Role of lncRNAs in the Nervous System

Several groups have examined what role if any, lncRNA play in the nervous system. One group identified 36 lncRNAs that were consistently downregulated when human ESCs (embryonic stem cells) were differentiated to NPCs (neural progenitor cells). Of these lncRNAs, three were found to be critical for differentiation of ESCs to NPCs, two of which, carry out this neurogenic role through interactions with REST and SUZ12 [84]. Another group performed an shRNA screen for lincRNAs that play a role in maintaining mESCs in a pluripotent state. From this initial screen, they identified TUNA (*Tcl1* Upstream Neuron-Associated lincRNA, also named megamind) as

being a lincRNA that is critical for proper commitment to the neural lineage. TUNA's neurogenic role is mediated through interactions with hnRNP-K, NCL, and PTBP1 and recruiting this complex to the promoters of neural genes, including Sox2, and Fgf4 during mESC differentiation. TUNA is also highly conserved and has a homolog in zebrafish, that when knocked down causes impairments in locomotor function [85].

One of the first mechanistic description of a neural specific lncRNA, was EVF2. EVF2 is a Shh (sonic hedgehog) regulated lncRNA that is located in the highly conserved Dlx-5/6 enhancer [86]. EVF2 functions both in cis, by forming anti-sense complementarity with DLX6 and trans by regulating the localization of DLX and MECP2 at the Dlx5/6 promoter. Furthermore, mice with impaired production of EVF2 have fewer GABAergic interneurons in various regions of the brain [87]. PNKY is another example of a conserved lncRNA that plays a critical role in neurogenesis and neural differentiation. PNKY was first identified by RNA-seq of various mouse brain regions and was found to be specifically expressed in the SVZ (subventricular zone) of adult mice, a region of the brain that is rich with NSCs [88,89]. Interestingly, knockdown of PNKY in NSCs in both an *in vitro* system and *in vivo* in SVZ NSCs, demonstrated a dramatic effect on stem cell dynamics, with a large reduction in the NSC population and an increase in transit-amplifying neuronal precursor population. The control of neuronal stem cell fate governed by PNKY appears to be mediated through its association with the RNA splicing protein, PTBP1, and altering the expression and splicing of genes involved in controlling cell lineage phenotypes [90].

In addition to controlling cell fate decisions in neuronal precursors, lncRNAs also play a role in developing neuronal connections. BDNF-AS is an antisense lncRNA that partially overlaps (over 200 bp) the gene for the neurotrophin, BDNF (brain-derived neurotrophic factor) [91,92]. BDNF-AS regulates BDNF protein levels through RNA-RNA interactions and when this

interaction is disrupted, BDNF protein levels increase by as much as 7 fold. Blocking or degrading BDNF-AS *in vitro* and *in vivo* resulted in an increase in neuronal outgrowth, survival and proliferation [92]. Another well characterized example is BC1/BC200, a lncRNA first identified in rats that is expressed almost entirely in the brain [93,94]. Following transcription, BC1/BC200 is actively transported from the neuron soma to the dendrites, due to a 62 nucleotide signal sequence in the 5' region of BC1/BC200 [95]. Once in the dendrites, BC1/BC200 decreases local translation by interacting with components of the translational machinery, eIF4a and PABP (polyA binding protein) [96]. The ultimate outcome of this translational inhibition is local control of synapse turnover and loss of BC1/BC200 in animal models leads to neuronal hyperexcitability and susceptibility to epileptogenic stress [96,97]. Furthermore, BC1/BC200 null mice have increased anxiety and display reduced exploratory behavior [98]. Taken together, these lncRNAs demonstrate the profound role that lncRNAs play in nervous system development and plasticity in zebrafish, mice and presumably humans.

lncRNAs in Noncancerous Brain Pathologies

In addition to being important in normal physiological processes, a growing body of work has provided evidence for lncRNAs being involved in several neural pathologies. Although there are some mechanistic studies implicating lncRNAs, much of the work associating lncRNAs to brain pathologies has been through GWAS (genome-wide association studies) and RNA-seq analysis. A GWAS analysis for SNPs associated with ASD (autism spectrum disorders) identified multiple variants near MSNP1AS (moesin pseudogene 1, anti-sense), a lncRNA that shares a large degree of sequence complementarity with MSN (moesin), a protein involved in neuronal architecture [99,100]. MSNP1AS interacts with MSN through direct RNA-RNA interactions and increases MSN protein level. Importantly, the GWAS hits were all associated with changes in

MSNP1AS expression levels, and MSNP1AS and MSN are found at higher levels in the brains of ASD patients, relative to normal brain controls [99]. A different example of a lncRNA involved in brain pathology is BACE1-AS. BACE1-AS is an evolutionarily conserved lncRNA that is antisense to BACE1 (β-secretase-1), an aspartic acid protease that is involved in cleavage of APP (amyloid precursor protein) leading to the production of amyloid-β 1-40 and by extension plays a direct role in the Alzheimer disease process [101]. BACE1-AS is upregulated in response to many cellular stresses, including amyloid-β cleavage products *in-vitro*. BACE1-AS expression levels are positively correlated with BACE-1 RNA and protein levels *in vitro* and *in vivo*. Interestingly, BACE1-AS stabilizes BACE1, through base pairing interactions, which spans a miR-485-5p target site on BACE1. By base pairing with BACE1, BACE1AS occludes the microRNA binding site and increases the half-life of BACE1 mRNA [102]. Furthermore, BACE1-AS is present at a higher abundance in Alzheimer's disease patient brains [103].

lncRNAs in Gliomas

LncRNAs are involved in many disease processes including gliomagenesis and recent studies have begun to show how lncRNAs are functioning in gliomas. HIF1A-AS2 (hypoxia-inducible factor 1 alpha-antisense RNA 2) is a lncRNA found to be highly expressed in mesenchymal subtype GSCs (GBM stem-like cells) compared to the other three GBM subtypes. Knockdown of HIF1A-AS2 caused a decrease in mesenchymal GSCs cell growth and neuroshpere formation *in vitro* and reduced xenograft tumor size in mice. Interestingly, HIF1A-AS2 is upregulated following hypoxic stress in mesenchymal GSCs but not proneural subtype GSCs. HIF1A-AS2 interacts with IGF2BP2 and DHX9 and this interaction is critical for the expression of HMGA1, a direct transcriptional target of both IGF2BP2 and DHX9. Another example of a lncRNA that plays an important role in gliomas is HULC (highly up-regulated in liver cancer).

HULC was first identified in a screen to identify deregulated genes in hepatocellular carcinoma [104]. Subsequent work has shown that HULC is dysregulated in a wide variety of tumors including esophageal carcinoma and pancreatic adenocarcinoma [105–108]. HULC is upregulated in gliomas, with GBM and grade III gliomas having the highest expression. HULC expression is positively correlated with angiogenesis associated factors ESM-1and VEGF, and HULC levels were positively correlated with microvessel density in patient tumors [105]. HULC knockdown has dramatic effects on many oncogenic phenotypes, including decrease in cell growth, colony formation, invasion adhesion and migration. HULC seems to alter the PI3K/AKT/mTOR signaling pathway through changing the levels of TGF-β1, although there is no clear mechanistic understanding on how HULC is actually doing this [105].

Several tumor suppressive lncRNAs have also been identified that play a role in glioma development. MEG3 is an imprinted, maternally expressed, lncRNA that is highly expressed in normal brain [109,110]. MEG3 is commonly downregulated in gliomas and its down-regulation is associated with worse patient outcomes. Reestablishing MEG3 expression in glioma cell lines results in a decrease in cell proliferation and increased apoptotic cell death, which is partly due to activating the p53 pathway by suppressing MDM2 [111]. Moreover, MEG3 knockout mice show over vascularization in fetal brains, suggesting that MEG3 downregulation in gliomas may not only function to increase cell growth but may also have pro-angiogenic functions [112]. GAS5 is another example of a tumor suppressive lncRNA in gliomas. GAS5 is strongly downregulated in GBM cell lines and overexpression of GAS5 results in cell death due to increases in bmf (Bcl-2-modifying factor) and subsequent activation of apoptosis. Interestingly, unlike previous reports of GAS5 acting as DNA-element decoy, in this context, GAS5 acts as a ceRNA by titrating mir-222, which directly targets bmf [52,113].

Since the early 80's when lncRNAs were first discovered, scientists have steadily made progress in increasing our understanding of their expression and function. Indeed, detailed descriptions of lncRNAs have broadened our understanding of cellular regulation and physiology. While progress has been made in many fields of lncRNA biology, relatively little progress has been made in identifying and characterizing lncRNAs in gliomas. Indeed, much work is still needed to fully catalog the lncRNAs that are deregulated in gliomas and to functionally characterize those associated with patient survival. This thesis will detail our work to comprehensively identify deregulated lncRNAs in gliomas and begin to functionally understand how a subset of these lncRNAs play a role in glioma tumor biology.

References

1. Ostrom QT, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, et al. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. Neuro-Oncology [Internet]. 2015 Oct 1;17(suppl 4):iv1-iv62. Available from: http://neuro-oncology.oxfordjournals.org/content/17/suppl_4/iv1.short

2. Bradl M, Lassmann H. Oligodendrocytes: biology and pathology. Acta Neuropathol [Internet]. 2010;119(1):37–53. Available from: http://dx.doi.org/10.1007/s00401-009-0601-5

3. Clarke LE, Barres BA. Emerging roles of astrocytes in neural circuit development. Nat Rev Neurosci [Internet]. 2013 May;14(5):311–21. Available from: http://dx.doi.org/10.1038/nrn3484

4. Filiano AJ, Gadani SP, Kipnis J. Interactions of innate and adaptive immunity in brain development and function. Brain Res [Internet]. 2015 Aug 18;1617:18–27. Available from: http://www.sciencedirect.com/science/article/pii/S0006899314010233

5. Zong H, Verhaak RGW, Canoll P. The cellular origin for malignant glioma and prospects for clinical advancements. Expert Rev Mol Diagn [Internet]. 2012 May 1;12(4):383–94. Available from: http://dx.doi.org/10.1586/erm.12.30

6. Rousseau A, Nutt CL, Betensky RA, Iafrate AJ, Han M, Ligon KL, et al. Expression of Oligodendroglial and Astrocytic Lineage Markers in Diffuse Gliomas: Use of YKL-40, ApoE, ASCL1, and NKX2-2. J Neuropathol &amp; Exp Neurol [Internet]. 2006 Dec 1;65(12):1149 LP-1156. Available from: http://jnen.oxfordjournals.org/content/65/12/1149.abstract

7.   Shoshan Y, Nishiyama A, Chang A, Mörk S, Barnett GH, Cowell JK, et al. Expression of oligodendrocyte progenitor cell antigens by gliomas: Implications for the histogenesis of brain tumors. Proc Natl Acad Sci [Internet]. 1999 Aug 31;96(18):10361–6. Available from: http://www.pnas.org/content/96/18/10361.abstract

8.   Gupta M, Djalilvand A, Brat DJ. Clarifying the Diffuse Gliomas. Am J Clin Pathol [Internet]. 2005 Nov 1;124(5):755 LP-768. Available from: http://ajcp.oxfordjournals.org/content/124/5/755.abstract

9.   Alcantara Llaguno S, Chen J, Kwon C-H, Jackson EL, Li Y, Burns DK, et al. Malignant Astrocytomas Originate from Neural Stem/Progenitor Cells in a Somatic Tumor Suppressor Mouse Model. Cancer Cell [Internet]. 2009 Jan 6;15(1):45–56. Available from: http://www.sciencedirect.com/science/article/pii/S1535610808004091

10.   Zhu X, Bergles DE, Nishiyama A. NG2 cells generate both oligodendrocytes and gray matter astrocytes. Development [Internet]. 2007 Dec 7;135(1):145 LP-157. Available from: http://dev.biologists.org/content/135/1/145.abstract

11.   Liu C, Sage JC, Miller MR, Verhaak RGW, Hippenmeyer S, Vogel H, et al. Mosaic Analysis with Double Markers Reveals Tumor Cell of Origin in Glioma. Cell [Internet]. 2011 Dec 1;146(2):209–21. Available from: http://dx.doi.org/10.1016/j.cell.2011.06.014

12.   SM C, IF P, Huang W, Al E. Patterns of care for adults with newly diagnosed malignant glioma. JAMA [Internet]. 2005 Feb 2;293(5):557–64. Available from: http://dx.doi.org/10.1001/jama.293.5.557

13.   Wheen LC, Anderson NE, Baker PCH, Singh VK, Synek BJL. Leptomeningeal infiltration as the presenting manifestation of a malignant glioma. J Clin Neurosci [Internet]. 2006 Feb;13(2):298–301. Available from: http://www.sciencedirect.com/science/article/pii/S0967586805004157

14.   Schüz J, Böhler E, Berg G, Schlehofer B, Hettinger I, Schlaefer K, et al. Cellular Phones, Cordless Phones, and the Risks of Glioma and Meningioma (Interphone Study Group, Germany). Am J Epidemiol [Internet]. 2006 Mar 15;163(6):512–20. Available from: http://aje.oxfordjournals.org/content/163/6/512.abstract

15.   Hardell L, Carlberg M, Hansson Mild K. Pooled analysis of two case–control studies on use of cellular and cordless telephones and the risk for malignant brain tumours diagnosed in 1997–2003. Int Arch Occup Environ Health [Internet]. 2006;79(8):630–9. Available from: http://dx.doi.org/10.1007/s00420-006-0088-5

16.   Lahkola A, Auvinen A, Raitanen J, Schoemaker MJ, Christensen HC, Feychting M, et al. Mobile phone use and risk of glioma in 5 North European countries. Int J Cancer [Internet]. 2007 Apr 15;120(8):1769–75. Available from: http://dx.doi.org/10.1002/ijc.22503

17.   Fisher JL, Schwartzbaum JA, Wrensch M, Wiemels JL. Epidemiology of Brain Tumors. Neurol Clin [Internet]. 2007 Nov;25(4):867–90. Available from:

http://www.sciencedirect.com/science/article/pii/S0733861907000746

18.   Wrensch M, Minn Y, Chew T, Bondy M, Berger MS. Epidemiology of primary brain tumors: Current concepts and review of the literature. Neuro-Oncology [Internet]. 2002 Oct 1;4(4):278–99. Available from: http://neuro-oncology.oxfordjournals.org/content/4/4/278.abstract

19.   Shintani T, Hayakawa N, Hoshi M, Sumida M, Kurisu K, Oki S, et al. High Incidence of Meningioma among Hiroshima Atomic Bomb Survivors. J Radiat Res [Internet]. 1999 Mar 1;40(1):49–57. Available from: http://jrr.oxfordjournals.org/content/40/1/49.abstract

20.   Preston-Martin S. EPIDEMIOLOGY OF PRIMARY CNS NEOPLASMS. Neurol Clin [Internet]. 1996 Dec 3;14(2):273–90. Available from: http://dx.doi.org/10.1016/S0733-8619(05)70256-5

21.   Brenner A V, Linet MS, Fine HA, Shapiro WR, Selker RG, Black PM, et al. History of allergies and autoimmune diseases and risk of brain tumors in adults . Int J Cancer [Internet]. 2002 May 10;99(2):252–9. Available from: http://dx.doi.org/10.1002/ijc.10320

22.   Wiemels JL, Wiencke JK, Patoka J, Moghadassi M, Chew T, McMillan A, et al. Reduced Immunoglobulin E and Allergy among Adults with Glioma Compared with Controls. Cancer Res [Internet]. 2004 Nov 17;64(22):8468 LP-8473. Available from: http://cancerres.aacrjournals.org/content/64/22/8468.abstract

23.   Wiemels JL, Wiencke JK, Sison JD, Miike R, McMillan A, Wrensch M. History of allergies among adults with glioma and controls . Int J Cancer [Internet]. 2002 Apr 1;98(4):609–15. Available from: http://dx.doi.org/10.1002/ijc.10239

24.   Wrensch M, Wiencke JK, Wiemels J, Miike R, Patoka J, Moghadassi M, et al. Serum IgE, Tumor Epidermal Growth Factor Receptor Expression, and Inherited Polymorphisms Associated with Glioma Survival. Cancer Res [Internet]. 2006 Apr 17;66(8):4531 LP-4541. Available from: http://cancerres.aacrjournals.org/content/66/8/4531.abstract

25.   Bondy M, Wiencke J, Wrensch M, Kyritsis AP. Genetics of primary brain tumors: a review. J Neurooncol [Internet]. 1994;18(1):69–81. Available from: http://dx.doi.org/10.1007/BF01324606

26.   Wrensch M, Lee M, Miike R, Newman B, Bargar G, Davis R, et al. Familial and Personal Medical History of Cancer and Nervous System Conditions among Adults with Glioma and Controls. Am J Epidemiol [Internet]. 1997 Apr 1;145(7):581–93. Available from: http://aje.oxfordjournals.org/content/145/7/581.abstract

27.   Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA Repair Genes and Associations with Cancer Risk. Cancer Epidemiol Biomarkers &amp; Prev [Internet]. 2002 Dec 1;11(12):1513 LP-1530. Available from: http://cebp.aacrjournals.org/content/11/12/1513.abstract

28.   Bond GL, Hu W, Bond EE, Robins H, Lutzker SG, Arva NC, et al. A Single Nucleotide

Polymorphism in the MDM2 Promoter Attenuates the p53 Tumor Suppressor Pathway and Accelerates Tumor Formation in Humans. Cell [Internet]. 2004 Nov 24;119(5):591–602. Available from: http://www.sciencedirect.com/science/article/pii/S0092867404010517

29. Wrensch M, Kelsey KT, Liu M, Miike R, Moghadassi M, Sison JD, et al. ERCC1 and ERCC2 polymorphisms and adult glioma. Neuro-Oncology [Internet]. 2005 Oct 1;7(4):495–507. Available from: http://neuro-oncology.oxfordjournals.org/content/7/4/495.abstract

30. Wang L-E, Bondy ML, Shen H, El-Zein R, Aldape K, Cao Y, et al. Polymorphisms of DNA Repair Genes and Risk of Glioma. Cancer Res [Internet]. 2004 Aug 16;64(16):5560 LP-5563. Available from: http://cancerres.aacrjournals.org/content/64/16/5560.abstract

31. Network CGAR. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. N Engl J Med [Internet]. 2015 Jun 10;372(26):2481–98. Available from: http://dx.doi.org/10.1056/NEJMoa1402121

32. Losman J-A, Kaelin WG. What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. Genes Dev [Internet]. 2013 Apr 15;27(8):836–52. Available from: http://genesdev.cshlp.org/content/27/8/836.abstract

33. Suzuki H, Aoki K, Chiba K, Sato Y, Shiozawa Y, Shiraishi Y, et al. Mutational landscape and clonal architecture in grade II and III gliomas. Nat Genet [Internet]. 2015 May;47(5):458–68. Available from: http://dx.doi.org/10.1038/ng.3273

34. Thakkar JP, Dolecek TA, Horbinski C, Ostrom QT, Lightner DD, Barnholtz-Sloan JS, et al. Epidemiologic and Molecular Prognostic Review of Glioblastoma. Cancer Epidemiol Biomarkers &amp; Prev [Internet]. 2014 Sep 30;23(10):1985 LP-1996. Available from: http://cebp.aacrjournals.org/content/23/10/1985.abstract

35. Ohgaki H, Kleihues P. Epidemiology and etiology of gliomas. Acta Neuropathol [Internet]. 2005;109(1):93–108. Available from: http://dx.doi.org/10.1007/s00401-005-0991-y

36. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell [Internet]. 2010 Dec 20;17(1):98–110. Available from: http://dx.doi.org/10.1016/j.ccr.2009.12.020

37. Laperriere N, Zuraw L, Cairncross G. Radiotherapy for newly diagnosed malignant glioma in adults: a systematic review. Radiother Oncol [Internet]. 2002 Sep;64(3):259–73. Available from: http://www.sciencedirect.com/science/article/pii/S0167814002000786

38. Mirimanoff R-O, Gorlia T, Mason W, Van den Bent MJ, Kortmann R-D, Fisher B, et al. Radiotherapy and Temozolomide for Newly Diagnosed Glioblastoma: Recursive Partitioning Analysis of the EORTC 26981/22981-NCIC CE3 Phase III Randomized Trial. J Clin Oncol [Internet]. 2006 Jun 1;24(16):2563–9. Available from:

http://ascopubs.org/doi/abs/10.1200/JCO.2005.04.5963

39.   Hegi ME, Diserens A-C, Gorlia T, Hamou M-F, de Tribolet N, Weller M, et al. MGMT
      Gene Silencing and Benefit from Temozolomide in Glioblastoma. N Engl J Med
      [Internet]. 2005 Mar 10;352(10):997–1003. Available from:
      http://dx.doi.org/10.1056/NEJMoa043331

40.   Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJB, Janzer RC, et al.
      Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy
      alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the
      EORTC-NCIC trial. Lancet Oncol [Internet]. 2009 May;10(5):459–66. Available from:
      http://www.sciencedirect.com/science/article/pii/S1470204509700257

41.   Stupp R, Taillibert S, AA K, al et. Maintenance therapy with tumor-treating fields plus
      temozolomide vs temozolomide alone for glioblastoma: A randomized clinical trial.
      JAMA [Internet]. 2015 Dec 15;314(23):2535–43. Available from:
      http://dx.doi.org/10.1001/jama.2015.16669

42.   Westphal M, Hilt DC, Bortey E, Delavault P, Olivares R, Warnke PC, et al. A phase 3
      trial of local chemotherapy with biodegradable carmustine (BCNU) wafers (Gliadel
      wafers) in patients with primary malignant glioma. Neuro-Oncology [Internet]. 2003 Apr
      1;5(2):79–88. Available from: http://neuro-
      oncology.oxfordjournals.org/content/5/2/79.abstract

43.   Lai A, Tran A, Nghiemphu PL, Pope WB, Solis OE, Selch M, et al. Phase II Study of
      Bevacizumab Plus Temozolomide During and After Radiation Therapy for Patients With
      Newly Diagnosed Glioblastoma Multiforme. J Clin Oncol [Internet]. 2011 Jan
      10;29(2):142–8. Available from: http://ascopubs.org/doi/abs/10.1200/JCO.2010.30.2729

44.   Vredenburgh JJ, Desjardins A, Reardon DA, Peters KB, Herndon JE, Marcello J, et al.
      The Addition of Bevacizumab to Standard Radiation Therapy and Temozolomide
      Followed by Bevacizumab, Temozolomide, and Irinotecan for Newly Diagnosed
      Glioblastoma. Clin Cancer Res [Internet]. 2011 Jun 15;17(12):4119 LP-4124. Available
      from: http://clincancerres.aacrjournals.org/content/17/12/4119.abstract

45.   Gilbert MR, Dignam JJ, Armstrong TS, Wefel JS, Blumenthal DT, Vogelbaum MA, et al.
      A Randomized Trial of Bevacizumab for Newly Diagnosed Glioblastoma. N Engl J Med
      [Internet]. 2014 Feb 19;370(8):699–708. Available from:
      http://dx.doi.org/10.1056/NEJMoa1308573

46.   Chinot OL, Wick W, Mason W, Henriksson R, Saran F, Nishikawa R, et al. Bevacizumab
      plus Radiotherapy–Temozolomide for Newly Diagnosed Glioblastoma. N Engl J Med
      [Internet]. 2014 Feb 19;370(8):709–22. Available from:
      http://dx.doi.org/10.1056/NEJMoa1308345

47.   Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. Ribosome Profiling
      Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. Cell [Internet].
      2013 Jul 3;154(1):240–51. Available from:

http://www.sciencedirect.com/science/article/pii/S0092867413007113

48. Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res . 2013 Apr;41(6):e74–e74.

49. McHugh CA, Chen C-K, Chow A, Surka CF, Tran C, McDonel P, et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature [Internet]. 2015 May 14;521(7551):232–6. Available from: http://dx.doi.org/10.1038/nature14443

50. Gupta R a, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature [Internet]. 2010;464(7291):1071–6. Available from: http://dx.doi.org/10.1038/nature08975

51. Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, et al. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature [Internet]. 2013 Aug 29;500(7464):598–602. Available from: http://dx.doi.org/10.1038/nature12451

52. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. Noncoding RNA Gas5 Is a Growth Arrest– and Starvation-Associated Repressor of the Glucocorticoid Receptor. Sci Signal [Internet]. 2010 Feb 2;3(107):ra8 LP-ra8. Available from: http://stke.sciencemag.org/content/3/107/ra8.abstract

53. Gonzalez I, Munita R, Agirre E, Dittmer TA, Gysling K, Misteli T, et al. A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. Nat Struct Mol Biol [Internet]. 2015 May;22(5):370–6. Available from: http://dx.doi.org/10.1038/nsmb.3005

54. Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat Struct Mol Biol. 2014 Feb;21(2):198–206.

55. Coccia EM, Cicala C, Charlesworth A, Ciccarelli C, Rossi GB, Philipson L, et al. Regulation and expression of a growth arrest-specific gene (gas5) during growth, differentiation, and development. Mol Cell Biol [Internet]. 1992 Aug;12(8):3514–21. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC364604/

56. Greenstein S, Ghias K, Krett NL, Rosen ST. Mechanisms of Glucocorticoid-mediated Apoptosis in Hematological Malignancies. Clin Cancer Res [Internet]. 2002 Jun 1;8(6):1681 LP-1694. Available from: http://clincancerres.aacrjournals.org/content/8/6/1681.abstract

57. Sun L, Goff LA, Trapnell C, Alexander R, Lo KA, Hacisuleyman E, et al. Long noncoding RNAs regulate adipogenesis. Proc Natl Acad Sci [Internet]. 2013 Feb 26;110(9):3387–92. Available from: http://www.pnas.org/content/110/9/3387.abstract

58. Wang J, Liu X, Wu H, Ni P, Gu Z, Qiao Y, et al. CREB up-regulates long non-coding

RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res [Internet]. 2010 Sep 1;38(16):5366–83. Available from: http://nar.oxfordjournals.org/content/38/16/5366.abstract

59.   Liu B, Sun L, Liu Q, Gong C, Yao Y, Lv X, et al. A Cytoplasmic NF-κB Interacting Long Noncoding RNA Blocks IκB Phosphorylation and Suppresses Breast Cancer Metastasis. Cancer Cell [Internet]. 2015 Mar 9;27(3):370–81. Available from: http://www.sciencedirect.com/science/article/pii/S1535610815000549

60.   Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K, et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature. 2013 Jan;493(7431):231–5.

61.   Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature [Internet]. 2012;491(7424):454–7. Available from: http://dx.doi.org/10.1038/nature11508

62.   Carrieri C, Forrest ARR, Santoro C, Persichetti F, Carninci P, Zucchelli S, et al. Expression analysis of the long non-coding RNA antisense to Uchl1 (AS Uchl1) during dopaminergic cells' differentiation in vitro and in neurochemical models of Parkinson's disease   [Internet]. Vol. 9, Frontiers in Cellular Neuroscience . 2015. p. 114. Available from: http://journal.frontiersin.org/article/10.3389/fncel.2015.00114

63.   Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, et al. A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response. Cell [Internet]. 2010 Aug 6;142(3):409–19. Available from: http://www.sciencedirect.com/science/article/pii/S0092867410007300

64.   Yoon J-H, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, et al. LincRNA-p21 Suppresses Target mRNA Translation. Mol Cell. 2012 Aug;47(4):648–55.

65.   Lai F, Gardini A, Zhang A, Shiekhattar R. Integrator mediates the biogenesis of enhancer RNAs. Nature [Internet]. 2015 Sep 17;525(7569):399–403. Available from: http://dx.doi.org/10.1038/nature14906

66.   Kaikkonen MU, Spann N, Heinz S, Romanoski CE, Allison KA, Stender JD, et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol Cell [Internet]. 2013 Aug 8;51(3):310–25. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3779836/

67.   Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature [Internet]. 2013 Jun 27;498(7455):516–20. Available from: http://dx.doi.org/10.1038/nature12210

68.   Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature [Internet]. 2013 Feb 28;494(7438):497–501. Available from:

http://dx.doi.org/10.1038/nature11884

69.   Mueller AC, Cichewicz MA, Dey BK, Layer R, Reon BJ, Gagan JR, et al. MUNC, a Long
       Noncoding RNA That Facilitates the Function of MyoD in Skeletal Myogenesis. Mol Cell
       Biol [Internet]. 2015 Feb 17;35(3):498–513. Available from:
       http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4285423/

70.   Mousavi K, Zare H, Dell'Orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, et al. ERNAs
       Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci.
       Mol Cell. 2013;

71.   Hsieh C-L, Fei T, Chen Y, Li T, Gao Y, Wang X, et al. Enhancer RNAs participate in
       androgen receptor-driven looping that selectively enhances gene activation. Proc Natl
       Acad Sci [Internet]. 2014 May 20;111(20):7319–24. Available from:
       http://www.pnas.org/content/111/20/7319.abstract

72.   Li W, Notani D, Rosenfeld MG. Enhancers as non-coding RNA transcription units: recent
       insights and future perspectives. Nat Rev Genet [Internet]. 2016 Apr;17(4):207–23.
       Available from: http://dx.doi.org/10.1038/nrg.2016.4

73.   Pasman Z, Been MD, Garcia-Blanco MA. Exon circularization in mammalian nuclear
       extracts. RNA [Internet]. 1996 Jun 1;2(6):603–10. Available from:
       http://rnajournal.cshlp.org/content/2/6/603.abstract

74.   Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, Oliner JD, et al. Scrambled exons.
       Cell [Internet]. 1991 Dec 8;64(3):607–13. Available from: http://dx.doi.org/10.1016/0092-
       8674(91)90244-S

75.   Cocquerelle C, Daubersies P, Majérus MA, Kerckaert JP, Bailleul B. Splicing with
       inverted order of exons occurs proximal to large introns. EMBO J [Internet]. 1992
       Mar;11(3):1095–8. Available from:
       http://www.ncbi.nlm.nih.gov/pmc/articles/PMC556550/

76.   Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are
       a large class of animal RNAs with regulatory potency. Nature [Internet]. 2013 Mar
       21;495(7441):333–8. Available from: http://dx.doi.org/10.1038/nature11928

77.   Zhang X-O, Wang H-B, Zhang Y, Lu X, Chen L-L, Yang L. Complementary Sequence-
       Mediated Exon Circularization. Cell [Internet]. 2014 Sep 25;159(1):134–47. Available
       from: http://www.sciencedirect.com/science/article/pii/S0092867414011118

78.   Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs Are the
       Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types.
       PLoS One [Internet]. 2012 Feb 1;7(2):e30733. Available from:
       http://dx.doi.org/10.1371%2Fjournal.pone.0030733

79.   Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-Type Specific Features of
       Circular RNA Expression. PLOS Genet [Internet]. 2013 Sep 5;9(9):e1003777. Available

from: http://dx.doi.org/10.1371%2Fjournal.pgen.1003777

80.   Conn SJ, Pillman KA, Toubia J, Conn VM, Salmanidis M, Phillips CA, et al. The RNA
       Binding Protein Quaking Regulates Formation of circRNAs. Cell [Internet]. 2015 Mar
       12;160(6):1125–34. Available from:
       http://www.sciencedirect.com/science/article/pii/S0092867415001798

81.   Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of
       mammalian circular RNAs. Genome Biol [Internet]. 2014;15(7):409. Available from:
       http://dx.doi.org/10.1186/s13059-014-0409-z

82.   Starke S, Jost I, Rossbach O, Schneider T, Schreiner S, Hung L-H, et al. Exon
       Circularization Requires Canonical Splice Signals. Cell Rep [Internet]. 2015 Dec
       8;10(1):103–11. Available from: http://dx.doi.org/10.1016/j.celrep.2014.12.002

83.   Zhang Y, Zhang X-O, Chen T, Xiang J-F, Yin Q-F, Xing Y-H, et al. Circular Intronic
       Long Noncoding RNAs. Mol Cell [Internet]. 2013 Sep 26;51(6):792–806. Available from:
       http://www.sciencedirect.com/science/article/pii/S109727651300590X

84.   Ng S, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and
       neuronal differentiation by association with chromatin modifiers and transcription factors.
       EMBO J [Internet]. 2011 Dec 23;31(3):522 LP-533. Available from:
       http://emboj.embopress.org/content/31/3/522.abstract

85.   Lin N, Chang K-Y, Li Z, Gates K, Rana ZA, Dang J, et al. An Evolutionarily Conserved
       Long Noncoding RNA TUNA Controls Pluripotency and Neural Lineage Commitment.
       Mol Cell [Internet]. 2014 Mar 20;53(6):1005–19. Available from:
       http://www.sciencedirect.com/science/article/pii/S1097276514000823

86.   Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD. The Evf-2 noncoding RNA is
       transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2
       transcriptional coactivator. Genes Dev [Internet]. 2006 Jun 1;20(11):1470–84. Available
       from: http://genesdev.cshlp.org/content/20/11/1470.abstract

87.   Bond AM, VanGompel MJW, Sametsky EA, Clark MF, Savage JC, Disterhoft JF, et al.
       Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal
       GABA circuitry. Nat Neurosci [Internet]. 2009 Aug;12(8):1020–7. Available from:
       http://dx.doi.org/10.1038/nn.2371

88.   Ramos AD, Diaz A, Nellore A, Delgado RN, Park K-Y, Gonzales-Roybal G, et al.
       Integration of Genome-wide Approaches Identifies lncRNAs of Adult Neural Stem Cells
       and Their Progeny In Vivo. Cell Stem Cell [Internet]. 2013 May 2;12(5):616–28.
       Available from: http://www.sciencedirect.com/science/article/pii/S1934590913000982

89.   Kriegstein A, Alvarez-Buylla A. The Glial Nature of Embryonic and Adult Neural Stem
       Cells. Annu Rev Neurosci [Internet]. 2009 Jun 1;32(1):149–84. Available from:
       http://dx.doi.org/10.1146/annurev.neuro.051508.135600

90.	Ramos AD, Andersen RE, Liu SJ, Nowakowski TJ, Hong SJ, Gertz CC, et al. The Long Noncoding RNA Pnky Regulates Neuronal Differentiation of Embryonic and Postnatal Neural Stem Cells. Cell Stem Cell. 2015 Apr;16(4):439–47.

91.	Liu Q-R, Walther D, Drgon T, Polesskaya O, Lesnick TG, Strain KJ, et al. Human brain derived neurotrophic factor (BDNF) genes, splicing patterns, and assessments of associations with substance abuse and Parkinson's Disease. Am J Med Genet Part B Neuropsychiatr Genet [Internet]. 2005 Apr 5;134B(1):93–103. Available from: http://dx.doi.org/10.1002/ajmg.b.30109

92.	Modarresi F, Faghihi MA, Lopez-Toledano MA, Fatemi RP, Magistri M, Brothers SP, et al. Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. Nat Biotech [Internet]. 2012 May;30(5):453–9. Available from: http://dx.doi.org/10.1038/nbt.2158

93.	Tiedge H, Fremeau RT, Weinstock PH, Arancio O, Brosius J. Dendritic location of neural BC1 RNA. Proc Natl Acad Sci [Internet]. 1991 Mar 15;88(6):2093–7. Available from: http://www.pnas.org/content/88/6/2093.abstract

94.	DeChiara TM, Brosius J. Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. Proc Natl Acad Sci [Internet]. 1987 May 1;84(9):2624–8. Available from: http://www.pnas.org/content/84/9/2624.abstract

95.	Muslimov IA, Santi E, Homel P, Perini S, Higgins D, Tiedge H. RNA Transport in Dendrites: A cis-Acting Targeting Element Is Contained within Neuronal BC1 RNA. J Neurosci [Internet]. 1997 Jun 15;17(12):4722 LP-4733. Available from: http://www.jneurosci.org/content/17/12/4722.abstract

96.	Wang H, Iacoangeli A, Popp S, Muslimov IA, Imataka H, Sonenberg N, et al. Dendritic BC1 RNA: Functional Role in Regulation of Translation Initiation. J Neurosci [Internet]. 2002 Dec 1;22(23):10232 LP-10241. Available from: http://www.jneurosci.org/content/22/23/10232.abstract

97.	Zhong J, Chuang S-C, Bianchi R, Zhao W, Lee H, Fenton AA, et al. BC1 Regulation of Metabotropic Glutamate Receptor-Mediated Neuronal Excitability. J Neurosci [Internet]. 2009 Aug 12;29(32):9977 LP-9986. Available from: http://www.jneurosci.org/content/29/32/9977.abstract

98.	Lewejohann L, Skryabin B V, Sachser N, Prehn C, Heiduschka P, Thanos S, et al. Role of a neuronal small non-messenger RNA: behavioural alterations in BC1 RNA-deleted mice. Behav Brain Res [Internet]. 2004 Sep 23;154(1):273–89. Available from: http://www.sciencedirect.com/science/article/pii/S0166432804000580

99.	Kerin T, Ramanathan A, Rivas K, Grepo N, Coetzee GA, Campbell DB. A Noncoding RNA Antisense to Moesin at 5p14.1 in Autism. Sci Transl Med. 2012 Apr;4(128):128ra40-128ra40.

100.	Furutani Y, Matsuno H, Kawasaki M, Sasaki T, Mori K, Yoshihara Y. Interaction

between Telencephalin and ERM Family Proteins Mediates Dendritic Filopodia Formation. J Neurosci [Internet]. 2007 Aug 15;27(33):8866 LP-8876. Available from: http://www.jneurosci.org/content/27/33/8866.abstract

101.   Monaco S, Zanusso G, Rizzuto SM and N. Cerebral Amyloidoses: Molecular Pathways and Therapeutic Challenges [Internet]. Vol. 13, Current Medicinal Chemistry. 2006. p. 1903–13. Available from: http://www.eurekaselect.com/node/56217/article

102.   Faghihi MA, Zhang M, Huang J, Modarresi F, Van der Brug MP, Nalls MA, et al. Evidence for natural antisense transcript-mediated inhibition of microRNA function. Genome Biol [Internet]. 2010;11(5):R56. Available from: http://dx.doi.org/10.1186/gb-2010-11-5-r56

103.   Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, et al. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of [beta]-secretase. Nat Med [Internet]. 2008 Jul;14(7):723–30. Available from: http://dx.doi.org/10.1038/nm1784

104.   Panzitt K, Tschernatsch MMO, Guelly C, Moustafa T, Stradner M, Strohmaier HM, et al. Characterization of HULC, a Novel Gene With Striking Up-Regulation in Hepatocellular Carcinoma, as Noncoding RNA. Gastroenterology [Internet]. 2007 Jan;132(1):330–42. Available from: http://www.sciencedirect.com/science/article/pii/S0016508506017859

105.   Zhu Y, Zhang X, Qi L, Cai Y, Yang P, Xuan G, et al. HULC long noncoding RNA silencing suppresses angiogenesis by regulating ESM-1 via the PI3K/Akt/mTOR signaling pathway in human gliomas. Oncotarget; Vol 7, No 12 [Internet]. 2016;7(12):14429–40. Available from: http://www.impactjournals.com/oncotarget/index.php?journal=oncotarget&amp

106.   Kang M, Sang Y, Gu H, Zheng L, Wang L, Liu C, et al. Long noncoding RNAs POLR2E rs3787016 C/T and HULC rs7763881 A/C polymorphisms are associated with decreased risk of esophageal cancer. Tumor Biol [Internet]. 2015;36(8):6401–8. Available from: http://dx.doi.org/10.1007/s13277-015-3328-z

107.   Peng W, Gao W, Feng J. Long noncoding RNA HULC is a novel biomarker of poor prognosis in patients with pancreatic cancer. Med Oncol [Internet]. 2014;31(12):346. Available from: http://dx.doi.org/10.1007/s12032-014-0346-4

108.   Matouk IJ, Abbasi I, Hochberg A, Galun E, Dweik H, Akkawi M. Highly upregulated in liver cancer noncoding RNA is overexpressed in hepatic colorectal metastasis. Eur J Gastroenterol Hepatol [Internet]. 2009;21(6):688–92. Available from: http://journals.lww.com/eurojgh/Fulltext/2009/06000/Highly_upregulated_in_liver_cancer_noncoding_RNA.14.aspx

109.   Zhang X, Sun S, Pu JKS, Tsang ACO, Lee D, Man VOY, et al. Long non-coding RNA expression profiles predict clinical phenotypes in glioma. Neurobiol Dis [Internet]. 2012 Oct [cited 2015 Aug 11];48(1):1–8. Available from: http://www.sciencedirect.com/science/article/pii/S0969996112002124

110. Miyoshi N, Wagatsuma H, Wakana S, Shiroishi T, Nomura M, Aisaka K, et al. Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q . Genes to Cells [Internet]. 2000 Mar 1;5(3):211–20. Available from: http://dx.doi.org/10.1046/j.1365-2443.2000.00320.x

111. Wang P, Ren Z, Sun P. Overexpression of the long non-coding RNA MEG3 impairs in vitro glioma cell proliferation. J Cell Biochem [Internet]. 2012 Jun 1;113(6):1868–74. Available from: http://dx.doi.org/10.1002/jcb.24055

112. Gordon FE, Nutt CL, Cheunsuchon P, Nakayama Y, Provencher KA, Rice KA, et al. Increased Expression of Angiogenic Genes in the Brains of Mouse Meg3-Null Embryos. Endocrinology [Internet]. 2010 Jun 14;151(6):2443–52. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2875815/

113. Zhao X, Wang P, Liu J, Zheng J, Liu Y, Chen J, et al. Gas5 Exerts Tumor-suppressive Functions in Human Glioma Cells by Targeting miR-222 [Internet]. Vol. 23, Mol Ther. American Society of Gene & Cell Therapy; 2015. p. 1899–911. Available from: http://dx.doi.org/10.1038/mt.2015.170

Chapter 2

Adapted from: Expression of lncRNAs in Low-Grade Gliomas and Glioblastoma Multiforme: An In Silico Analysis

Reon BJ, Anaya, J, Zhang Y, Mandell J, Purow B, Abounader R, and Dutta A.

Expression of lncRNAs in Low-Grade Gliomas and Glioblastoma Multiforme: An In Silico Analysis

Brian J. Reon[1,2], Jordan Anaya[2], Ying Zhang[3], James Mandell[1], Benjamin Purow[3], Roger Abounader[3], Anindya Dutta[2]*

1 Department of Pathology, School of Medicine, University of Virginia, Charlottesville, Virginia, United States of America,

2 Department of Biochemistry, University of Virginia, Charlottesville, Virginia, United States of America,

3 Division of Neuro-Oncology, Neurology Department, University of Virginia Health System, Old Medical School, Charlottesville, Virginia, United States of America

Abstract

Background

Each year, over 16,000 patients die from malignant brain cancer in the US. Long noncoding RNAs (lncRNAs) have recently been shown to play critical roles in regulating neurogenesis and brain tumor progression. To better understand the role of lncRNAs in brain cancer, we performed a global analysis to identify and characterize all annotated and novel lncRNAs in both grade II and III gliomas as well as grade IV glioblastomas (glioblastoma multiforme [GBM]).

Methods and Findings

We determined the expression of all lncRNAs in over 650 brain cancer and 70 normal brain tissue RNA sequencing datasets from The Cancer Genome Atlas (TCGA) and other publicly available datasets. We identified 611 induced and 677 repressed lncRNAs in glial tumors relative to normal brains. Hundreds of lncRNAs were specifically expressed in each of the three lower

grade glioma (LGG) subtypes (IDH1/2 wt, IDH1/2 mut, and IDH1/2 mut 1p19q codeletion) and the four subtypes of GBMs (classical, mesenchymal, neural, and proneural). Overlap between the subtype-specific lncRNAs in GBMs and LGGs demonstrated similarities between mesenchymal GBMs and IDH1/2 wt LGGs, with 2-fold higher overlap than would be expected by random chance. Using a multivariate Cox regression survival model, we identified 584 and 282 lncRNAs that were associated with a poor and good prognosis, respectively, in GBM patients. We developed a survival algorithm for LGGs based on the expression of 64 lncRNAs that was associated with patient prognosis in a test set (hazard ratio [HR] = 2.168, 95% CI = 1.765–2.807, $p < 0.001$) and validation set (HR = 1.921, 95% CI = 1.333–2.767, $p < 0.001$) of patients from TCGA. The main limitations of this study are that further work is needed to investigate the clinical relevance of our findings, and that validation in an independent dataset is needed to determine the robustness of our survival algorithm.

Conclusions

This work identifies a panel of lncRNAs that appear to be prognostic in gliomas and provides a critical resource for future studies examining the role of lncRNAs in brain cancers.

Introduction

Malignant gliomas are the most common aggressive primary brain tumor, with nearly 23,000 new cases diagnosed each year in the US [1]. The most aggressive malignant gliomas, anaplastic astrocytoma and glioblastoma multiforme (GBM), have 5-y survival rates of 23% and 5%, respectively. World Health Organization grade II and III gliomas are less aggressive than grade IV glioblastomas (GBMs), and have been grouped together by The Cancer Genome Atlas (TCGA) as lower grade gliomas (LGGs). Once thought to be a single disease, GBMs are now recognized as having a considerable level of intertumor heterogeneity, and studies have found that GBMs can be subdivided into four subtypes, proneural, neural, classical, and mesenchymal, based on their transcriptional profile [2,3]. Importantly, these subtypes are associated with differing clinical outcomes, including varying responses to intensive therapy and differences in overall survival [3]. Similar to GBMs, LGGs can be categorized into distinct subtypes, IDH1/2 mut, IDH1/2 mut 1p19q codeletion, and IDH1/2 wt, based on IDH1/2 mutational status and the presence of a codeletion of 1p19q [4]. Each subtype has distinct clinical phenotypes, with the IDH1/2 wt subtype being the most aggressive and dissimilar to the other LGG subtypes [4,5]. Although knowledge of tumor subtype has clinical utility, the best prognostic indicator for patients with glial tumors is the mutational status of IDH1 and IDH2 [6]. In LGGs, patients with wild-type IDH1/2 have a median survival of 1.7 y, while those with mutant IDH1/2 have a median survival between 6.3 and 8.0 y. In GBMs, the corresponding median survival estimates are 1.1 and 2.1 y for wild-type and mutant IDH1/2, respectively [4].

Proteins have been thought to be the primary functional effectors of cells until relatively recently, when the roles of noncoding RNAs (ncRNAs) began to be appreciated for their contributions to most biological processes. Spurred by large sequencing consortia such as

ENCODE and FANTOM, interest in ncRNAs has grown rapidly, in part due to the discovery that the vast majority of the mammalian genome is transcribed and that most of the resulting transcripts do not code for proteins [7–9]. Long noncoding RNAs (lncRNAs) are a class of ncRNAs greater than 200 bp in length that do not code for a protein. lncRNAs have mechanistically diverse functions in the cell, and in the nucleus have been shown to regulate gene expression either in cis or in trans by recruiting chromatin-modifying complexes to promoters of target genes [10,11]. Also, lncRNAs have been found to regulate gene expression by promoting long-distance genomic interactions [12]. Other, mainly cytoplasmic, lncRNAs have been shown to regulate the protein concentrations produced from target genes in part by affecting mRNA stability or the translational efficiency of an mRNA [13–15].

Recent work has shown that lncRNAs play a critical role in various biological pathways including the immune system [16], muscle differentiation [17,18], neural lineage commitment, lineage specification, and synaptogenesis [19–23]. In addition to their role in normal physiological processes, lncRNAs are also important regulators of disease processes [24,25]. In cancer, lncRNAs can act as either tumor suppressors or oncogenes, and have been shown to regulate tumor growth and metastasis in breast, prostate, and liver cancer [10,26–28]. Although some lncRNAs have been linked to brain tumor development and pathogenesis, the overall study of lncRNAs in brain tumors has lagged behind [29–32]. In this study, we sought to categorize all dysregulated lncRNAs in glial tumors and to identify lncRNAs that are associated with patient prognosis.

Methods

Ethics Statement

All patient samples, including 15 primary GBM specimens and five normal brain specimens, were obtained from consented patients undergoing surgical treatment at the University of Virginia and were acquired in accordance with a protocol approved by the University of Virginia's institutional review board.

Planning the Analysis

There was no protocol or prospective analysis plan for the study. Unaligned sequencing reads for TCGA GBMs and LGGs were downloaded from the Cancer Genomics Hub. Normal brain sample SRA files (SRP033725, SRP045638, SRP044668, and SRP048683) [33–36] were downloaded from the Sequence Read Archive (SRA) database on October 4, 2014. Most RNA sequencing (RNA-seq) samples in TCGA originate from patients from the US; however, TCGA collects patient samples from other countries as well, including Canada, Russia, and Italy. The LGG survival algorithm was devised in November 2015; 70% of patients were used as a test set, and 30% were retained for a validation set. Following suggestions by a reviewer in September 2016, this division was subsequently changed to 60% of patients in the test set and the remaining 40% in the validation set.

Identification of Novel lncRNAs and Quantification of lncRNA Abundance

The aforementioned SRA files were converted to fastq files using the SRA Toolkit v2.3.5. All fastq files were aligned to the hg38 reference genome with Tophat2 using default settings [37]. Novel transcripts (transcripts not found in reference transcript annotation files: GENCODE and RefSeq) were identified in each sample using Cufflinks2 de novo assembly [38]. A consensus transcript assembly was generated using Cuffmerge. Novel transcripts whose genomic coordinates did not intersect with known transcripts from a custom GTF file consisting

of transcripts from GENCODE v21, RefSeq, and Cabili et al. [39] were kept for further validation.

We determined the coding potential for each novel transcript using CPAT (Coding-Potential Assessment Tool) and intersection with a mass spectrometry database. First, the in silico coding potential of each novel transcript was assessed using CPAT [40], and any transcripts with a CPAT score above 0.5242 were considered transcripts of unknown coding potential and were not included in downstream analysis. Second, we mapped all unique peptides from the ProteomicsDB [41] mass spectrometry database to all known proteins. All potential ORFs within each novel transcript were translated, and all unmapped peptides were mapped on the translated ORFs. Any novel transcript with more than one mapped peptide was not considered for downstream analysis. All novel transcripts that met these criteria were considered as novel lncRNAs and added to our finalized GTF file. Using the finalized GTF file, the expression of all genes was quantified using Cuffquant and Cuffnorm.

Validating lncRNA Expression in Clinical Samples

Fresh-frozen GBM and normal brain tissue samples were obtained from the University of Virginia, and RNA was isolated using Trizol (Thermo Fisher). RNA treated with DNase (RQ1 Promega, Thermo Fisher) was used for reverse transcription with SuperScript III (Thermo Fisher). Quantitative real-time PCR (RT-PCR) was performed on tissue cDNA using SYBR Green (Thermo Fisher), and the expression of LINC00152, TUNAR, and LINC01476 was normalized to that of the housekeeping gene encoding actin.

Identifying Differentially Expressed lncRNAs and lncRNAs Associated with Mutation Status and Subtype

To identify differentially expressed lncRNAs, we selected only lncRNAs with a median expression greater than 0.5 fragments per kilobase of exon per million fragments mapped (FPKM); 4,288 lncRNAs in LGGs and 3,297 lncRNAs in GBMs met this threshold and were used in our downstream analysis. Expression values for each lncRNA in 170 GBM samples, 497 LGG samples (we removed 16 samples from our initial pool of 513 LGGs due to >15% of transcripts having expression values three standard deviations above or below the mean expression for all LGGs), and 78 normal tissue samples were used to calculate the Kolmogorov-Smirnov test (KS test) statistic [42]. lncRNAs with a Benjamini-Hochberg-corrected false discovery rate (FDR) [43] of <0.05 and a fold change greater than or equal to four were considered differentially expressed. The correlation between copy number variation (CNV) and lncRNA dysregulation was determined by calculating the Spearman correlation coefficient between lncRNA expression and the copy number segment mean (Tier 3 TCGA data accessed from the Broad GDAC Firehose; https://gdac.broadinstitute.org) of the genomic region that gives rise to each lncRNA. lncRNAs with Spearman correlation coefficients greater than or equal to 0.2 were considered to be correlated. The overlap between GBM and LGG differentially expressed lncRNAs was measured using the Jaccard index, a statistic that compares the similarity and divergence of two datasets, defined as the intersection of two datasets divided by the union of the datasets: $J(A,B) = \dfrac{A \cap B}{A \cup B}$.

To identify lncRNAs associated with individual somatic mutations, we separated patients into two groups based on the presence of nonsynonymous mutation or a non-inframe insertion or deletion in commonly mutated protein-coding genes (prevalence of 5% or greater). Differential expression was measured using the same statistical method mentioned above, with a log2 fold change of greater than 0.5 as a cutoff.

Subtype-associated lncRNAs were selected by separating tumors into groups based on their subtype and using the KS test statistic to determine if the expression of a given lncRNA was different in a specific subtype (FDR < 0.05). Specifically, a lncRNA was considered subtype specific only if its expression was statistically different from each other subtype in a paired comparison. GBMs were separated into subtypes based on predesignated subtyping (Cancer Genome Browser). LGGs subtypes were determined by identifying the mutational status of IDH1 or IDH2 from TCGA's preprocessed mutation calling data and identifying LGGs with 1p19q deletions from TCGA's preprocessed CNV data. The statistical significance of the overlap between GBM and LGG subtype-specific lncRNAs was determined by comparing the observed lncRNA overlap to the overlap obtained through 1,000 random iterations of the two sample sets, keeping the number of lncRNAs in each sample equal to that in the observed subtype-specific lncRNA sets.

LGG Survival Prediction

To create a lncRNA survival model for LGGs, we first randomly selected 60% of patients to serve as a test set and reserved the remaining 40% of patients for independent validation. Random subsamples of 66% of the test set of patients were subjected to a multivariate Cox regression [44] survival model. This was repeated with 100 random subsamples from the test set of patients. Age, grade, sex, IDH1/2 mutational status, and inverse normalized lncRNA expression levels were used as variables in the LGG survival model. In total, 64 lncRNAs had statistically significant Cox coefficients in 80% of the 100 subsamples and were included in our survival algorithm. To combine the predictive power of the prognostic lncRNAs, the following steps were taken. For every patient in our test set, the expression of each prognostic lncRNA was compared to the average expression of that lncRNA in patients from the test set. If the absolute

value of the expression Z-score of a given lncRNA was ≥1, the median Cox coefficient of that

lncRNA from the 100 subsamples was added to a summed Cox coefficient, and this was repeated

for each of the prognostic lncRNAs (S2 and S3 Figs). Patients were divided into two groups

based on whether the summed Cox coefficient was positive (poor prognosis) or negative (good

prognosis), consistent with the interpretation that a Cox coefficient > 0 indicates a poor outcome

and a Cox coefficient < 0 indicates a good outcome. The survival differences of the groups were

displayed on a Kaplan-Meier survival curve. We independently applied this survival algorithm,

as mentioned above, to the validation patient set (the retained 40% of patients).

lncRNAs Associated with Survival in GBM and GBM Subtypes

Prognostic lncRNAs were identified using a Cox proportional hazard model similar to

that stated above, except that we included age, sex, and inverse normalized lncRNA expression

in the survival model. lncRNAs that were associated with prognosis (p-value < 0.05) were then

separated based on whether they predicted a poor or good prognosis. We predicted the possible

pathways that each lncRNA is involved in using guilt-by-association analysis, as previously

described [45]. To identify lncRNAs that predict survival in each subtype, we performed Cox

regression for each lncRNA in a given subtype (based on subtypes specified by the Cancer

Genome Browser) and selected only lncRNAs with a p-value of ≤0.05.

Results

Identifying Novel lncRNAs

In order to identify and catalog all novel lncRNAs (unannotated lncRNAs) in brain

cancers, we used the Tuxedo Suite [37,38] to align, assemble, and quantify the expression of

novel and annotated transcripts from 170 GBMs and 497 LGGs originating from TCGA and 78

normal brain samples from both TCGA and publicly available datasets (Fig 1A). We initially

filtered all novel transcripts from the consensus transcriptome that (1) overlapped with any annotated transcript, (2) were less than 200 bp, or (3) did not contain a splice junction. We next assessed the coding potential of all novel transcripts using both in silico predictions as well as intersection with a human proteome database with peptide data from over 100 cell lines and 60 tissues that, importantly, includes brain tissue [41].

CPAT (Coding-Potential Assessment Tool) determines the coding potential of a novel transcript based on relative ORF size, codon bias, and nucleotide hexamer bias. Any transcripts with a CPAT score ≥ 0.5242—a threshold that separates noncoding RNAs and protein-coding genes [46]—were removed from further consideration (Fig 1B). We next sought to determine if there was any biological evidence of protein products derived from the novel transcripts by parsing data from ProteomicsDB [41]. Peptides from the database were first aligned to all known proteins, and any unaligned peptide was then aligned to all translated ORFs within the novel transcripts. Any transcript with two or more mapped peptides, suggesting that it could potentially be a novel protein-coding gene, was removed from downstream analysis. Only seven novel transcripts had more than two aligned peptides and were removed from our downstream analysis, which is in line with other studies reporting low levels of spurious ribosomal associations with ncRNAs [47]. After filtering our list of novel transcripts, we identified over 2,700 novel lncRNAs. After cataloging the novel lncRNAs expressed in LGGs, GBMs, and normal brain, the analyses were carried out with the entire pool of lncRNAs (both novel and annotated) whose average expression was greater than 0.5 FPKM.

Similar to protein-coding genes, lncRNAs are often transcribed by RNA polymerase II and share similar active chromatin marks of H3K4me3 on their promoters. We therefore tested whether H3K4me3 was located at the promoters of the novel lncRNAs identified in this study,

using H3K4me3 ChIP-seq data from U87 cells. The promoters of the novel lncRNAs were enriched in H3K4me3 chromatin marks relative to a randomized genomic control, although to a lesser extent than the promoters of protein-coding genes (Fig 1C).

Dysregulation of lncRNAs in Brain Cancers

Although recent work has begun to address the role of lncRNAs in brain tumors, very few lncRNAs have been found to be dysregulated in glial tumors [30,48]. Therefore, we sought to form a comprehensive list of all lncRNAs whose expression is significantly altered in brain tumors. TCGA RNA-seq data for glial tumors have very few accompanying normal brain samples, making comparisons between normal and tumor groups difficult due to a lack of adequate sample size. To bolster our ability to identify dysregulated lncRNAs in glial tumors, we included RNA-seq data from publicly available normal brain samples that were obtained from regions of the brain where glial tumors commonly arise (e.g., cortex, and excluding regions such as hippocampus and cerebellum; see Methods).

Using our normal brain samples as a comparison, we tested whether any lncRNAs were over- or underexpressed in GBMs or LGGs relative to normal brain. We identified 454 upregulated and 642 downregulated lncRNAs in GBMs and 403 upregulated and 340 downregulated lncRNAs in LGGs that all had FDRs of <0.05, a statistic that takes into account errors that may arise from multiple testing (see Methods), and had fold changes greater than four. Of these dysregulated lncRNAs, over 80 were newly identified in this study. We tested whether the expression differences of the dysregulated lncRNAs could be explained by genomic CNV in GBMs and LGGs. Consistent with previous studies [49], only a fraction (19% and 20% from GBMs and LGGs, respectively) of the differentially expressed lncRNAs were associated (Spearman correlation coefficient of 0.2 or greater) with tumor CNV (S1 Fig). This suggests that

other mechanisms, in addition to CNV, play a role in regulating the changes in lncRNA expression observed in glial tumors.

We have highlighted several differentially expressed lncRNAs in GBMs and LGGs in Fig 2A and 2B, respectively. Of note, our analysis confirmed previous work that identified a lncRNA, CRNDE, that is upregulated in a number of tumors, including GBMs [29]; in our analysis, CRNDE is upregulated over 40-fold in GBMs compared to normal brain (Fig 2A). Furthermore, we also identified TUNAR as being severely downregulated in all glial tumors, almost 45-fold in GBMs and 14-fold in LGGs (Fig 2A and 2B). This is interesting, as other work has shown that TUNAR is a crucial positive regulator of neuronal development and differentiation in zebrafish, mice, and humans, which suggests that brain tumors require the downregulation of TUNAR in order to gain oncogenic properties and escape the restrictions on neuronal cell growth [50,51]. In order to further validate our analysis, we measured by RT-PCR the expression of one lncRNA, LINC00152, which is upregulated 20-fold in GBMs in TCGA data (Fig 2A). Using normal brain and GBM tumor tissue from patients at the University of Virginia, we validated the altered expression of three lncRNAs, LINC00152, TUNAR, and LINC01476. LINC00152 was found to have 3-fold higher expression in tumor tissue relative to normal brain tissue (Fig 2C). TUNAR and LINC01476, which both have lower expression in GBMs relative to normal brain tissue in TCGA, were found to have 12-fold and nearly 100-fold higher expression in normal brain tissue compared to GBM tissue, respectively (Fig 2C).

We next tested whether there is any overlap between the differentially expressed lncRNAs in GBMs and LGGs. Interestingly, there was a large degree of overlap in both the upregulated and downregulated lncRNAs, with a Jaccard index (described in Methods) of 0.4 and 0.45, respectively (Fig 2D). Unlike other tumors, whose tumor grades are more commonly

viewed as being along a disease continuum, GBMs and their grade II and III counterparts are not typically regarded as being different stages of a single disease [49]. However, our results suggest that there is a great deal of similarity in the lncRNA profile of GBMs and LGGs. Some of the overlap could be due to the need of glial tumors to downregulate genes related to the differentiation of glia or neurons, though it is unlikely that such de-differentiation would account for such a high degree of overlap between LGGs and GBMs.

lncRNAs Associated with Patient Tumor Mutation Status

Somatic mutations are well-known drivers of tumorigenesis, and their profound effects on the cell's transcriptional landscape have been well characterized [52–54]. Although most studies have focused on changes in protein-coding gene expression, recent work has begun to show that somatic mutations can lead to large alterations in lncRNA expression as well [55–57]. Through the TCGA consortium, many recurrent somatic mutations have been identified in GBMs and LGGs, many of which are shared between the two groups [4,58]. In order to determine what effect these mutations might have on the lncRNA transcriptome, for each commonly mutated gene, we separated patients into groups based on their tumor mutational status and then tested whether the expression of any lncRNA is altered in GBMs and LGGs harboring each common somatic mutation.

We identified hundreds of lncRNAs that were differentially expressed (as described in Methods) in mutated versus non-mutated GBMs and LGGs (Fig 3A and 3C). Interestingly, in GBMs there was little overlap in the differentially expressed lncRNAs between different somatic mutation groups (Fig 3B). In contrast, LGGs had a higher degree of overlap between mutation-associated lncRNAs (Fig 3D).

lncRNAs Associated with Cancer Subtypes

Work by TCGA and others has found that both GBMs and LGGs are not homogenous collections of tumors, but can rather be categorized into separate subtypes [2–5]. Each of the glial tumor subtypes is clinically distinct, and understanding the lncRNAs associated with a particular subtype could help to better distinguish between the groups or possibly identify novel therapeutic targets. To this end, we separated patients based on their tumor subtype and determined whether there were any lncRNAs that were specifically expressed in a given subtype. We identified 64, 211, 95, and 71 lncRNAs that were specifically up- or downregulated in neural, proneural, mesenchymal, and classical GBMs, respectively (Fig 4A). Thirteen of these lncRNAs were novel lncRNAs identified in this study. Furthermore, 1,357, 1,216, and 466 lncRNAs were specifically up- or downregulated in IDH1/2 wt, IDH1/2 mut, and IDH1/2 mut 1p19q codeletion LGGs, respectively (Fig 4B).

Traditionally, GBMs and LGGs have been viewed as distinct oncological entities; however, recent work has begun to suggest that IDH1/2 wt LGGs might be more similar to GBMs than to their IDH1/2 mut LGG counterparts [49]. In order to better understand these similarities, we tested whether there is significant overlap (as described in Methods) between the differentially expressed genes (DEGs) of the IDH1/2 wt LGGs and the DEGs for each GBM subtype. Although there was no statistically significant overlap in DEGs between the neural GBM subtype and the IDH1/2 wt LGG subtype, the proneural GBM subtype had much less overlap with the IDH1/2 wt LGG subtype than would be expected by random chance (Fig 4C). This finding is consistent with the fact that proneural GBMs frequently have point mutations in IDH1/2 [3]. There was a slight increase in the overlap between classical GBM subtype DEGs and IDH1/2 wt LGG DEGs compared to the random model; however, this difference was not statistically significant ($p = 0.055$). Surprisingly, DEGs in mesenchymal GBMs had much higher

overlap with DEGs in IDH1/2 wt LGGs compared to the random model (Fig 4C). We next

determined whether the overlap between mesenchymal GBMs and IDH1/2 wt LGGs is specific

to this LGG subtype or is found with the other LGG subtypes, by measuring the overlap of

mesenchymal differentially expressed lncRNAs with differentially expressed lncRNAs from

each LGG subtype. In contrast to the greater degree of overlap with the IDH1/2 wt subtype, both

the IDH1/2 mut and IDH1/2 mut 1p19q codeletion subtypes had less overlap than would be

expected by random chance (Fig 4D). These similarities in the lncRNA profiles of IDH1/2 wt

LGGs and mesenchymal GBMs suggest that LGGs with wild-type IDH1/2 may progress to

mesenchymal GBMs.

lncRNA Expression and Survival in LGG Patients

The main prognostic variable for patients with glial tumors is the mutational status of

IDH1 or IDH2. In LGGs, recent work has shown that patients whose tumors also harbor 1p19q

codeletions have a slightly better overall survival than patients with IDH1/2 mut tumors without

1p19q codeletions [4]. We decided to test whether the expression of lncRNAs can be used to

separate patients into distinct survival groups, independent of IDH1/2 mutational status. To this

end, we performed survival analysis utilizing a multivariate Cox proportional hazard model that

included IDH1/2 mutational state, age, sex, tumor grade, and lncRNA expression as independent

variables in the survival model. It is common to find extreme outliers in large RNA-seq datasets,

which can negatively impact survival regression analysis. In order to correct for these outliers,

the expression values for each lncRNA were inverse normal transformed, a procedure that

increases the sensitivity and specificity of regression analysis using RNA-seq expression values

[59]. To attempt to separate patients into distinct prognostic groups using lncRNAs, we

randomly assigned 60% of LGG patients with complete clinical data (269 patients) to a test set,

on which we performed Cox regression to identify lncRNAs associated with survival in this patient cohort. We then selected all lncRNAs that were significantly associated with survival and created a survival algorithm with variables for each lncRNA that were weighted based on each lncRNA's contribution to overall survival (see Methods). This survival algorithm was then applied to the remaining 40% of patients (180 patients) who constituted the validation set (Figs 5A, S2 and S3).

After performing Cox regression on our test set, we identified 64 lncRNAs that were consistently associated with survival. These lncRNAs were included in our survival algorithm, which was then applied to the test set. Each patient received a score, based on how many prognostic lncRNAs met our expression cutoff (see Methods), and patients were then divided into two groups, good prognosis and poor prognosis. Our algorithm separated the test set into groups of 85 and 184 patients with median survival times of 1,209 and 4,084 d, respectively (HR = 2.168, 95% CI = 1.765–2.807, $p < 0.001$) (Fig 5B). We next applied this survival algorithm to the validation set and were successfully able to separate patients into distinct groups of 66 and 114 patients with median survival times of 2,235 and 4,412 d, respectively (HR = 1.921, 95% CI = 1.333–2.767, $p < 0.001$) (Fig 5C).

Identifying lncRNAs Associated with Survival in GBMs

We next sought to identify all lncRNAs that were associated with overall survival in patients with GBMs. Using Cox regression we identified 584 lncRNAs that were associated with a poor prognosis and 282 lncRNAs that were associated with better survival outcomes. A subset of these lncRNAs were independently used to separate GBM patients based on lncRNA expression levels in the top third and bottom third of patients (55 patients), and Kaplan-Meier plots show that these groups were associated with prognosis with statistical significance (Fig 6A

). Patients with high expression of RP11-334C17.6 had a median survival time of 485 d, while patients with low expression had a median survival time of 380 d (HR = 0.728, 95% CI = 0.6011–0.883, p = 0.00122) (Fig 6A). Patients with high and low expression of BTAT10 had median survival times of 335 and 485 d, respectively (HR = 1.298, 95% CI = 1.0881–1.548, p = 0.00374) (Fig 6B). However, unlike in the LGGs, we have not yet succeeded in combining the individually predictive lncRNAs into a survival algorithm that can predict prognosis in GBMs with statistical significance.

Unlike for proteins, ascertaining the function of a lncRNA based on sequence composition is extremely difficult. However, studies have shown that it is possible to infer what biological pathways a lncRNA might function in using guilt-by-association analysis, a technique that infers association of a lncRNA with a biological pathway based on the pathway enrichment of protein-coding genes whose expression is highly correlated with the lncRNA [45]. We used guilt-by-association analysis to determine what biological pathways are enriched in our positive and negative prognostic lncRNA groups. Interestingly, lncRNAs that are associated with a better prognosis in GBMs are more likely to be associated with signaling pathways, showing enrichment in protein kinase and phosphorylation pathways as well as signal transduction pathways. Conversely, lncRNAs that are associated with poor patient outcomes are highly associated with cell cycle pathways, immune response, and chromosome organization (Table 1).

We next subdivided all of the GBM tumors into their respective subtypes and performed Cox regression with all lncRNAs in each subtype: 165, 128, 88, and 385 lncRNAs were associated with prognosis in classical, mesenchymal, neural, and proneural GBM subtypes, respectively. Given the transcriptional, clinical, and phenotypic differences between the subtypes, we then tested if there was any overlap in the identities of the positive and negative

prognostic lncRNAs between subtypes. There was very little overlap noted between the prognostic lncRNAs in each subtype (Fig 6C and 6D), consistent with the hypothesis that the GBM subtypes arise from distinct mutational backgrounds and have very different biology.

Discussion

Our analysis of RNA-seq data for grade II, III, and IV brain tumors and normal brain tissue has identified hundreds of dysregulated lncRNAs in glial tumors, many of which are associated with tumor subtype or mutational status. Using Cox regression, we identified a panel of 64 lncRNAs that are associated with survival in LGG patients. We also identified lncRNAs that are similarly associated with prognosis in each GBM subtype and found remarkably little overlap of prognostic lncRNAs between GBM subtypes.

The growing appreciation for the important roles that lncRNAs play in tumor development and progression necessitates having a means of prioritizing which lncRNAs should be studied in a given cancer type. Global analyses have been performed for tumor types other than GBMs and LGGs, such as squamous cell lung carcinomas and adenocarcinomas, as well as meta-analyses of all tumors within TCGA [46,57,60]. Although meta-analyses of lncRNAs are extremely important, they have not been especially informative for brain tumors for several reasons. First, due to the broad nature of the analyses, it is not possible to focus on the specific nuances of each tumor type (i.e., subtypes). Second, due to the limited number of normal brain samples in TCGA, GBMs and LGGs were not included in many of the meta-analyses, which relied on comparisons with a reference panel of normal tissues. Lastly, the effect of lncRNA expression on survival in individual tumor types was not a main focus of the studies [46,60]. This is important because, depending on the tumor context, a given lncRNA may act as a tumor suppressor or oncogene [61,62]. By focusing specifically on brain tumors and including over 70

normal brain tissue samples, our analysis provides unique insights into the roles of lncRNAs in aggressive brain cancers.

In addition to studying the roles of annotated lncRNAs, we identified 2,706 novel multi-exon lncRNAs that are present in either normal brain tissue or brain tumors, but are not annotated in the commonly used lncRNA databases (GENCODE and RefSeq). Many of these novel lncRNAs were differentially expressed in brain tumors and were associated with clinically relevant mutations. Although the exact mechanisms leading to altered lncRNA expression are not known, roughly 20% of differentially expressed lncRNAs were weakly correlated with chromosomal amplifications and deletions. We also identified several hundred lncRNAs that were associated with GBM and LGG subtypes. It is well known that IDH1/2 wt LGGs have clinical phenotypes and genomic alterations similar to those of primary GBMs [5]. Interestingly, the intersection of subtype-associated lncRNAs between GBMs and LGGs revealed transcriptional similarities between IDH1/2 wt LGGs and mesenchymal GBMs. Although our analysis suggests an evolutionary link between mesenchymal GBMs and IDH1/2 wt LGGs, it does not preclude other tumor evolutionary pathways leading to the formation of mesenchymal GBMs. Other groups have found evidence that suggests non-GCIMP (non-glioma-CpG island methylator phenotype) mesenchymal GBMs evolve from a proneural GBM precursor [63]. However, this evolutionary pathway does not explain the origin of all mesenchymal GBMs. Our analysis suggests that some mesenchymal GBMs might arise from undetected IDH1/2 wt LGGs, which at clinical presentation would appear to be mesenchymal GBMs.

There are several limitations to this analysis. One limitation is that, because RNA-seq data from TCGA were derived from bulk tumor specimens, we are unable to decipher whether the differences in expression that we have identified are a reflection of alterations in tumor cells'

transcriptional programs or a reflection of tumor heterogeneity and changes in the stromal composition of each individual tumor. Furthermore, although we validated the expression changes of some lncRNAs using independent patient-derived samples, more work is needed to confirm the expression differences we identified for lncRNAs between gliomas and normal brain, and among tumor subtypes and tumors of different mutational status. Another limitation is that the generation and testing of our survival algorithm were performed on the same dataset. Although the "validation" dataset was blinded in the algorithm generation, further validation of our survival analysis in a truly independent dataset is needed to determine the significance and robustness of the survival algorithm. Independent validation of lncRNAs that are associated with survival in GBMs and GBM subtypes is also needed.

As stated earlier, IDH1/2 mutational status is the primary prognostic indicator for glial tumors. Using multivariate survival analysis, we have shown that a panel of lncRNAs can be used to separate LGG patients into distinct prognostic groups. This group of lncRNAs could potentially be used to help identify at-risk patients who might require more intensive therapies, although further validation in an independent dataset is needed to fully test the utility of the survival algorithm. Furthermore, we have also found several hundred lncRNAs that are prognostic in GBMs as a whole, as well as in individual subtypes. In summary, we have performed the first global analysis, to our knowledge, of lncRNAs in LGGs and GBMs. Our analysis can serve as a valuable resource for those working in the field to prioritize which lncRNAs to study in brain cancers.

Acknowledgments

References

1. Ostrom QT, Gittleman H, Liao P, Rouse C, Chen Y, Dowling J, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. Neuro Oncol. 2014 Oct;16(Suppl 4):iv1-63.

2. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, 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):157-73.

3. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010;17(1):98-110.

4. Cancer Genome Atlas Research Network, Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. N Engl J Med. 2015;372(26):2481-98.

5. Suzuki H, Aoki K, Chiba K, Sato Y, Shiozawa Y, Shiraishi Y, et al. Mutational landscape and clonal architecture in grade II and III gliomas. Nat Genet. 2015;47(5):458-68.

6. van den Bent MJ, Dubbink HJ, Marie Y, Brandes AA, Taphoorn MJB, Wesseling P, et al. IDH1 and IDH2 mutations are prognostic but not predictive for outcome in anaplastic oligodendroglial tumors: a report of the European Organization for Research and Treatment of Cancer Brain Tumor Group. Clin Cancer Res. 2010;16(5):1597-604.

7. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447(7146):799-816.

8. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. Science. 2005;309(5740):1559-63.

9. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. Science. 2005;309(5740):1564-6.

10. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071-6.

11. Trimarchi T, Bilal E, Ntziachristos P, Fabbri G, Dalla-Favera R, Tsirigos A, et al. Genome-wide mapping and characterization of notch-regulated long noncoding RNAs in acute leukemia. Cell. 2014;158(3):593-606.

12. Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat Struct Mol Biol. 2014;21(2):198-206.

13. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature. 2012;491(7424):454-7.

14. Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K, et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature. 2013;493(7431):231-5.

15. Yoon J-H, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, et al. LincRNA-p21 suppresses target mRNA translation. Mol Cell. 2012;47(4):648-55.

16. Gomez JA, Wapinski OL, Yang YW, Bureau J-F, Gopinath S, Monack DM, et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-γ locus. Cell. 2016;152(4):743-54.

17. Mousavi K, Zare H, Dell'Orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, et al. ERNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell. 2013;51(5):606-17.

18. Mueller AC, Cichewicz MA, Dey BK, Layer R, Reon BJ, Gagan JR, et al. MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis. Mol Cell Biol. 2015;35(3):498-513.

19. Ng SY, Johnson, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. EMBO J. 2012;31(3):522-33.

20. Ramos AD, Andersen RE, Liu SJ, Nowakowski TJ, Hong SJ, Gertz CC, et al. The long noncoding RNA Pnky regulates neuronal differentiation of embryonic and postnatal neural stem cells. Cell Stem Cell. 2015;16(4):439-47.

21. Rapicavoli NA, Poth EM, Blackshaw S. The long noncoding RNA RNCR2 directs mouse retinal cell specification. BMC Dev Biol. 2010;10(1):1-10.

22. Aprea J, Prenninger S, Dori M, Ghosh T, Monasor LS, Wessendorf E, et al. Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. EMBO J. 2013;32(24):3145-60.

23. Bernard D, Prasanth K V, Tripathi V, Colasse S, Nakamura T, Xuan Z, et al. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. EMBO J. 2010;29(18):3082-93.

24. Kerin T, Ramanathan A, Rivas K, Grepo N, Coetzee GA, Campbell DB. A noncoding RNA antisense to moesin at 5p14.1 in autism. Sci Transl Med. 2012;4(128):128ra40.

25. Zhao X, Tang Z, Zhang H, Atianjoh FE, Zhao J-Y, Liang L, et al. A long noncoding RNA contributes to neuropathic pain by silencing Kcna2 in primary afferent neurons. Nat Neurosci. 2013;16(8):1024-31.

26. Yuan JH, Yang F, Wang F, Ma JZ, Guo YJ, Tao QF, et al. A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. Cancer Cell. 2014;25(5):666-81.

27. Sakurai K, Reon BJ, Anaya J, Dutta A. The lncRNA DRAIC/PCAT29 locus constitutes a tumor-suppressive nexus. Mol Cancer Res. 2015;13(5):828-38.

28. Malik R, Patel L, Prensner JR, Shi Y, Iyer M, Subramaniyan S, et al. The lncRNA PCAT29 inhibits oncogenic phenotypes in prostate cancer. Mol Cancer Res. 2014;12(8):1081-7.

29. Ellis BC, Molloy PL, Graham LD. CRNDE: a long non-coding RNA involved in cancer, neurobiology and development. Front Genet. 2012;3:270.

30. Wang Y, Wang Y, Li J, Zhang Y, Yin H, Han B. CRNDE, a long-noncoding RNA, promotes glioma cell growth and invasion through mTOR signaling. Cancer Lett. 2015;367(2):122-8.

31. Zhang K, Sun X, Zhou X, Han L, Chen L. Long non-coding RNA HOTAIR promotes glioblastoma cell cycle progression in an EZH2 dependent manner. Oncotarget. 2014;6(1):537-46.

32. Zhang X, Kiang KM, Zhang GP, Leung GK. long non-coding RNAs dysregulation and function in glioblastoma stem cells. Noncoding RNA. 2015;1(1):69-86.

33. Gill BJ, Pisapia DJ, Malone HR, Goldstein H, Lei L, Sonabend A, et al. MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. Proc Natl Acad Sci U S A. 2014;111(34):12550-5.

34. Akula N, Barb J, Jiang X, Wendland JR, Choi KH, Sen SK, et al. RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. Mol Psychiatry. 2014;19(11):1179-85.

35. Jaffe AE, Shin J, Collado-Torres L, Leek JT, Tao R, Li C, et al. Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. Nat Neurosci. 2015;18(1):154-61.

36. Li J, Shi M, Ma Z, Zhao S, Euskirchen G, Ziskin J, et al. Integrated systems analysis reveals a molecular network underlying autism spectrum disorders. Mol Syst Biol. 2014;10(12):774.

37. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg S. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36.

38. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotech. 2013;31(1):46-53.

39. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011;25(18):1915-27.

40. Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013;41(6):e74.

41. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, et al. Mass-spectrometry-based draft of the human proteome. Nature. 2014;509(7502):582-7.

42. Massey FJ Jr. The Kolmogorov-Smirnov test for goodness of fit. J Am Stat Assoc. 1951;46(253):68-78.

43. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57(1):289-300.

44. Cox DR. Regression models and life-tables. J R Stat Soc Ser B. 1972;34(2):187-220.

45. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458(7235):223-7.

46. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015;47(3):199-208.

47. Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, et al. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife. 2013;2013(2):1-24.

48. Shi Y, Wang Y, Luan W, Wang P, Tao T, Zhang J, et al. Long non-coding RNA H19 promotes glioma cell invasion by deriving miR-675. PLoS ONE. 2014;9(1):e86295.

49. Wen PY, Kesari S. Malignant gliomas in adults. N Engl J Med. 2008;359(5):492-507.

50. Lin N, Chang KY, Li Z, Gates K, Rana Z, Dang J, et al. An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. Mol Cell. 2014;53(6):1005-19.

51. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell. 2011;147(7):1537-50.

52. Gerstung M, Pellagatti A, Malcovati L, Giagounidis A, Porta MG, Jädersten M, et al. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. Nat Commun. 2015;6:5901.

53. Stead LF, Berri S, Wood HM, Egan P, Conway C, Daly C, et al. The transcriptional consequences of somatic amplifications, deletions, and rearrangements in a human lung squamous cell carcinoma. Neoplasia. 2012;14(11):1075-86.

54. Bashashati A, Haffari G, Ding J, Ha G, Lui K, Rosner J, et al. DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. Genome Biol. 2012;13(12):R124.

55. Garzon R, Volinia S, Papaioannou D, Nicolet D, Kohlschmidt J, Yan PS, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc Natl Acad Sci U S A. 2014;111(52):18679-84.

56. Flockhart RJ, Webster DE, Qu K, Mascarenhas N, Kovalski J, Kretz M, et al. BRAFV600E remodels the melanocyte transcriptome and induces BANCR to regulate melanoma cell migration. Genome Res. 2012;22(6):1006-14.

57. White N, Cabanski C, Silva-Fisher J, Dang H, Govindan R, Maher C. Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. Genome Biol. 2014;15(8):429.

58. Brennan CW, Verhaak RGW, McKenna A, Campos B, Noushmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. Cell. 2013;155(2):462-77.

59. Zwiener I, Frisch B, Binder H. Transforming RNA-Seq data to improve the performance of prognostic gene signatures. PLoS ONE. 2014;9(1):e85150.

60. Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, et al. Comprehensive genomic characterization of long non-coding RNAs across human cancers. Cancer Cell. 2015;28(4):529-40.

61. Matouk IJ, DeGroot N, Mezan S, Ayesh S, Abu-lail R, Hochberg A, et al. The H19 non-coding RNA is essential for human tumor growth. PLoS ONE. 2007;2(9):e845.

62. Yoshimizu T, Miroglio A, Ripoche M-A, Gabory A, Vernucci M, Riccio A, et al. The H19 locus acts in vivo as a tumor suppressor. Proc Natl Acad Sci U S A. 2008;105(34):12417-22.

63. Ozawa T, Riester M, Cheng Y-K, Huse JT, Squatrito M, Helmy K, et al. Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. Cancer Cell. 2014;26(2):288-300.
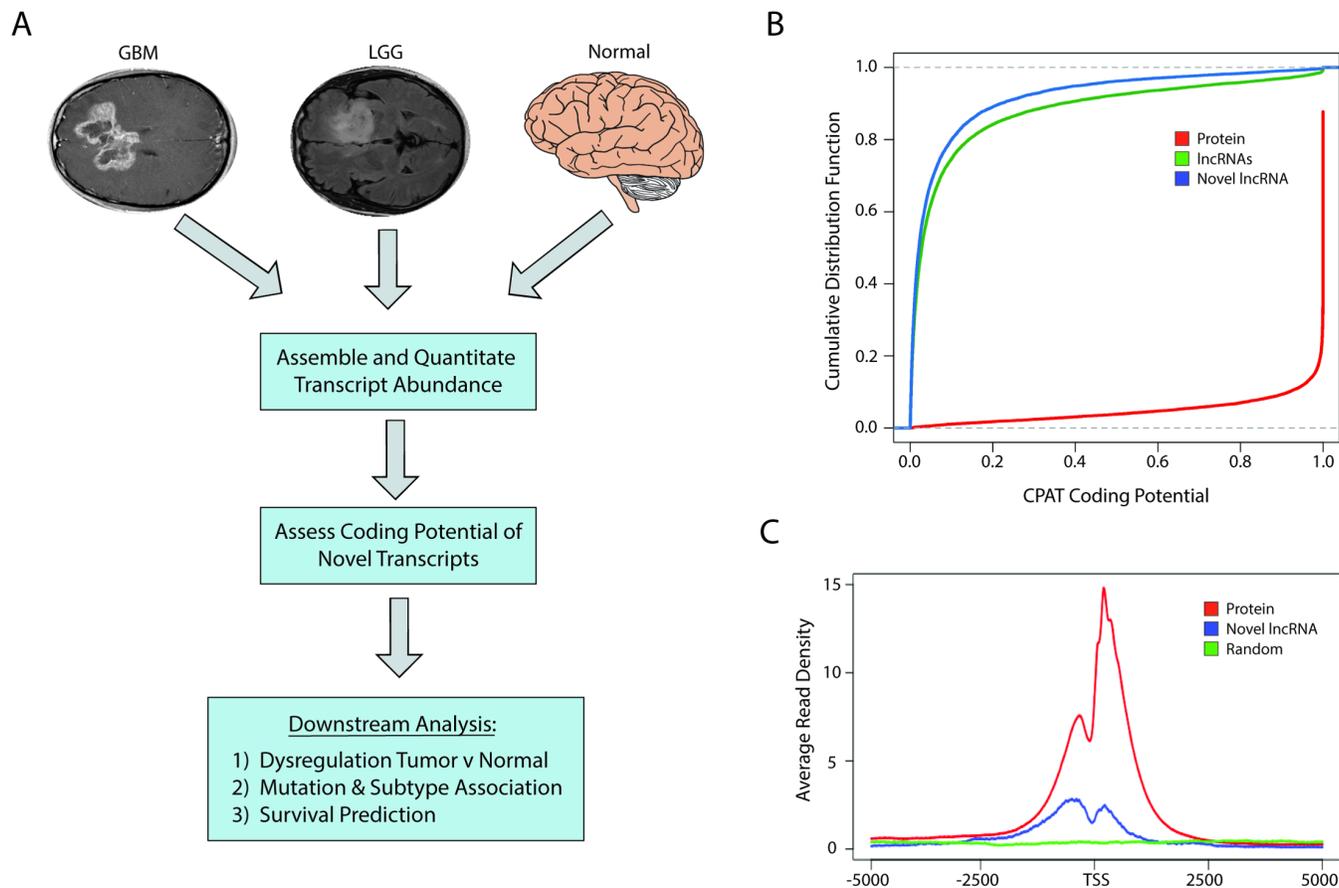
Fig 1. Identification of novel lncRNAs in glial tumors and normal brain RNA-seq samples.
(A) Overview of analysis pipeline for identifying novel lncRNAs and determining their associations with clinically relevant phenotypes. (B) Cumulative distribution function plot of CPAT scores demonstrates that the majority of novel transcripts are predicted to not code for proteins (CPAT score < 0.5242). (C) Metagene plot of H3K4me3 ChIP-seq data from U87 cell samples shows enrichment in promoters of protein-coding genes and novel lncRNAs but not in a randomized genomic control. GBM, glioblastoma multiforme; LGG, lower grade glioma; lncRNA, long noncoding RNA; RNA-seq, RNA sequencing.
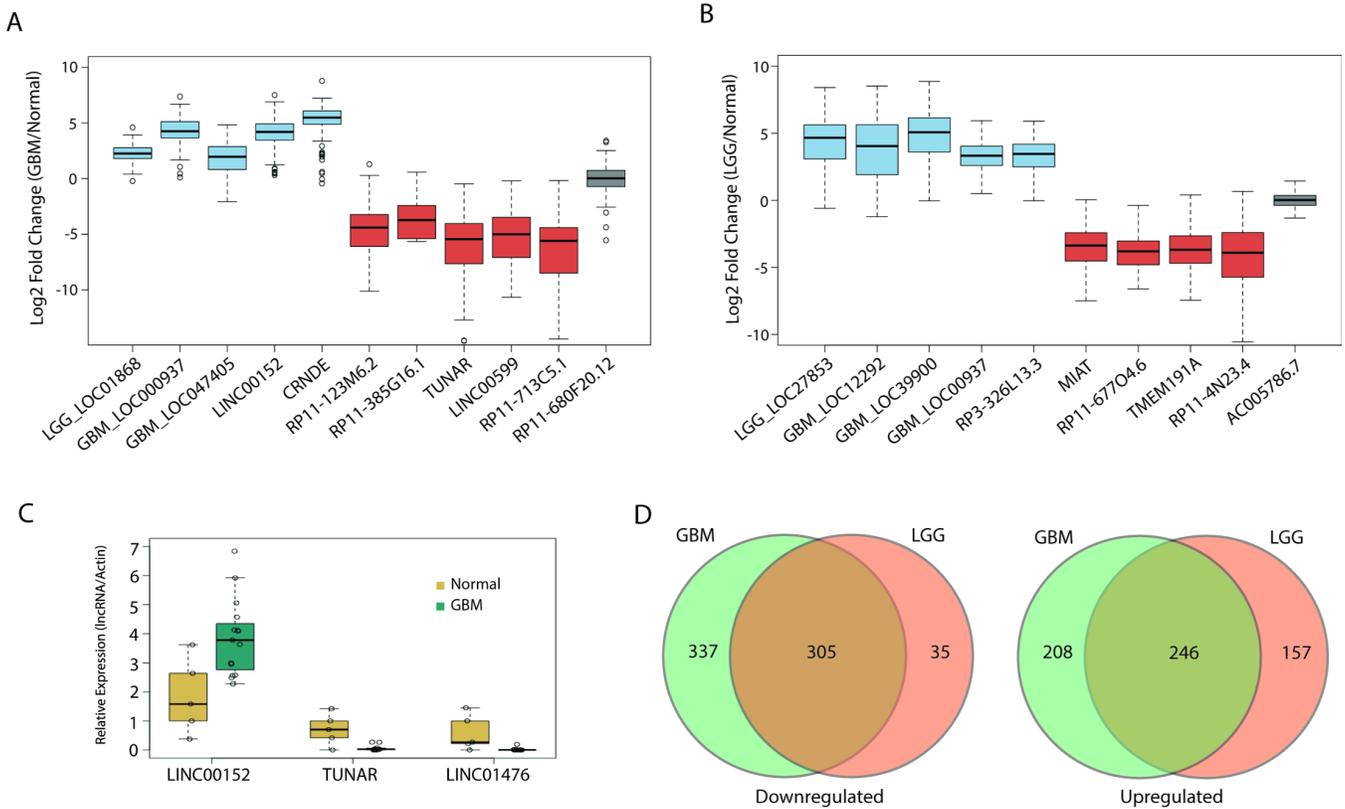
Fig 2. Large alterations in lncRNA expression in glial tumors.
(A) Boxplot of ten candidate lncRNAs that are differentially expressed in GBMs compared to normal tissue (blue = upregulated, red = downregulated, grey = no change). (B) Boxplot of nine candidate lncRNAs that are differentially expressed in LGGs compared to normal tissue (blue = upregulated, red = downregulated, grey = no change). (C) Real-time PCR of LINC00152, LINC01476, and TUNAR in 15 GBM and five normal brain samples confirms upregulation of LINC00152 and downregulation of LINC01476 and TUNAR in GBMs compared to normal brain. Expression values for each lncRNA are normalized to that of the gene encoding actin. (D) Large overlap in differentially expressed lncRNAs between GBMs and LGGs. GBM, glioblastoma multiforme; LGG, lower grade glioma; lncRNA, long noncoding RNA.
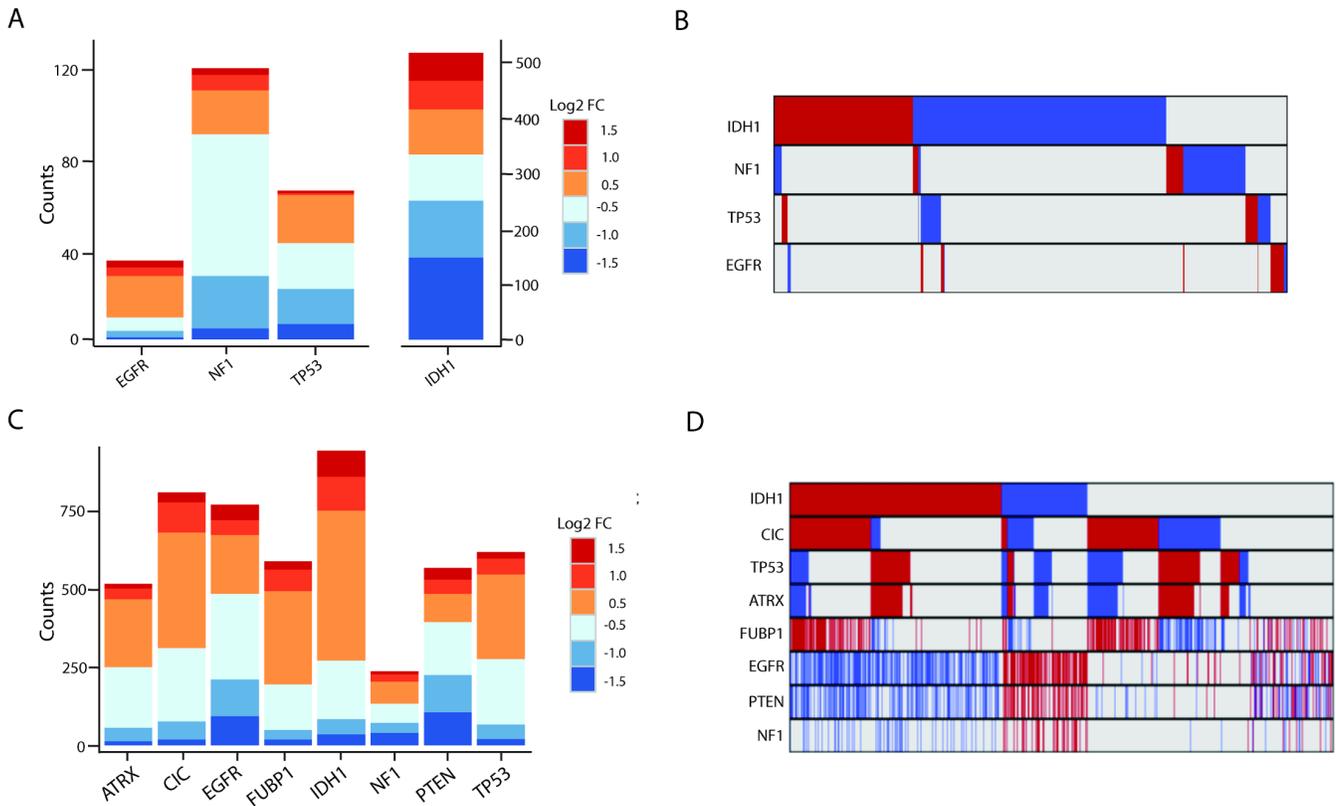
Fig 3. Many lncRNAs are associated with common mutations in GBMs and LGGs.
(A) Stacked bar graph of mutation-associated lncRNAs in GBMs. (B) Minimal overlap between
mutation-associated lncRNAs in GBMs (red = upregulated, blue = downregulated, grey = no change).
(C) Stacked bar graph of mutation-associated lncRNAs in LGGs shows robust deregulation depending
on tumor mutational background. (D) Moderate overlap between mutation-associated lncRNA
expression trends in GBMs; however, each group of mutation-associated lncRNAs represents a
distinct set of GBMs (red = upregulated, blue = downregulated, grey = no change). GBM,
glioblastoma multiforme; LGG, lower grade glioma; lncRNA, long noncoding RNA.
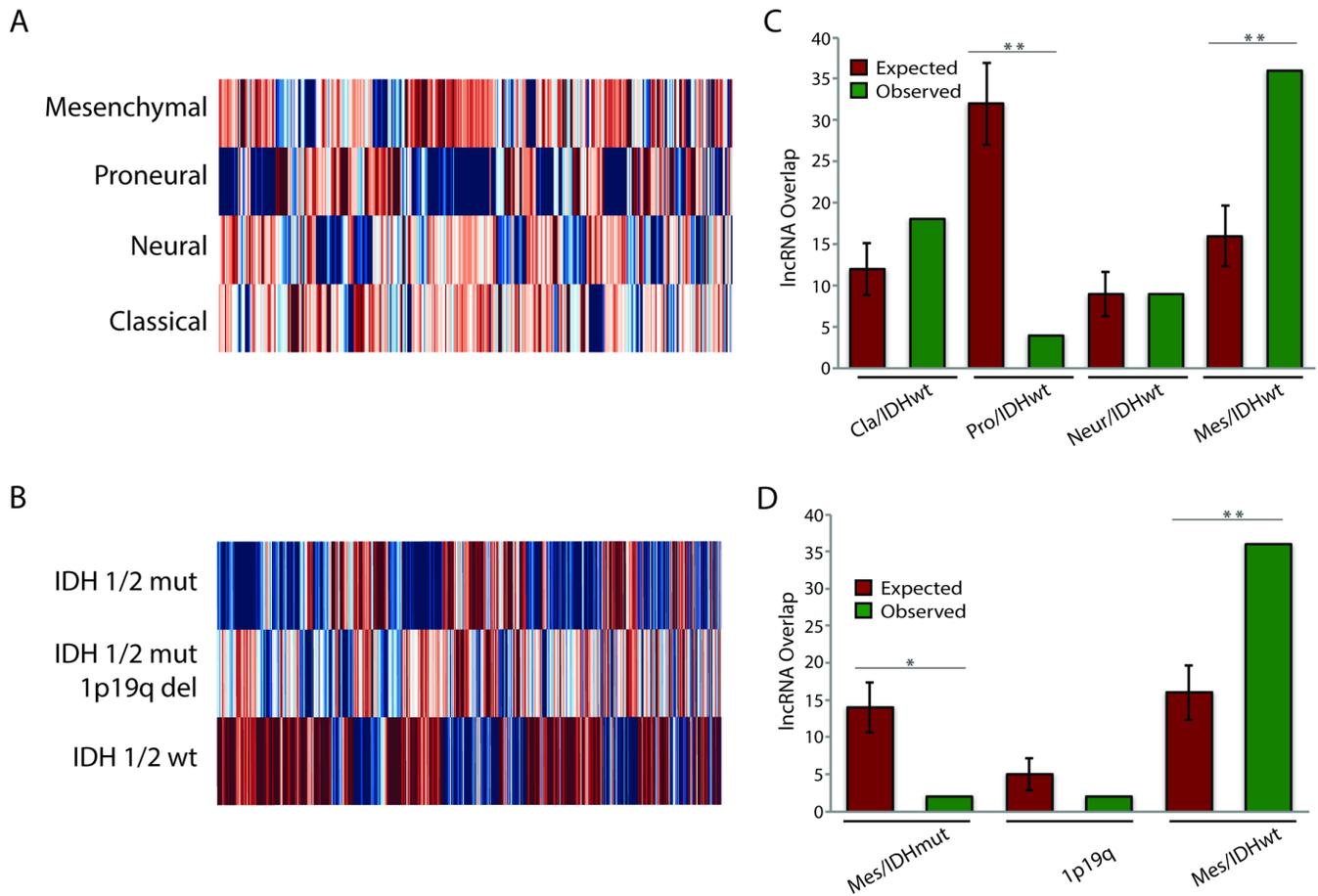
Fig 4. Identification of subtype-associated lncRNAs reveals similarities between GBM and LGG subtypes.
(A) Heatmap of all lncRNAs in GBMs with subtype-specific expression patterns demonstrates large expression differences between GBM subtypes. (B) Heatmap of all lncRNAs in LGGs with subtype-specific expression patterns demonstrates large expression differences between LGG subtypes. (C) Overlap of IDH1/2 wt LGG subtype-specific lncRNAs with GBM subtype-specific lncRNAs reveals similarities between mesenchymal GBMs and IDH1/2 wt LGGs. (D) Overlap of GBM mesenchymal subtype-specific lncRNAs with each group of LGG subtype-specific lncRNAs reveals similarities between mesenchymal GBMs and IDH1/2 wt LGGs. * $p < 0.00005$, ** $p < 0.00001$. GBM, glioblastoma multiforme; LGG, lower grade glioma; lncRNA, long noncoding RNA.
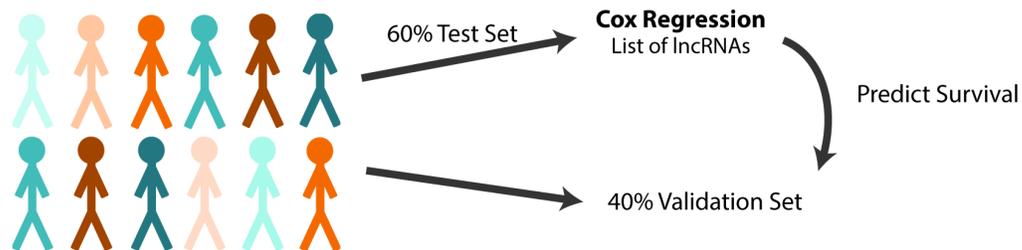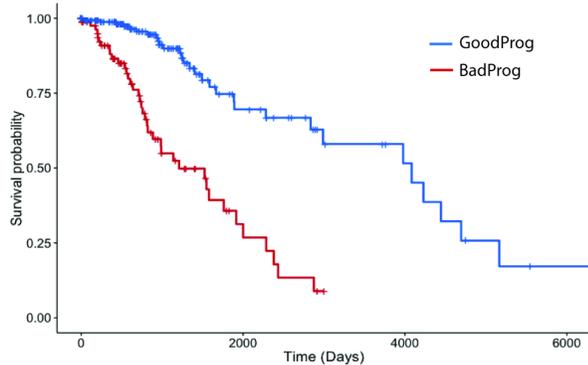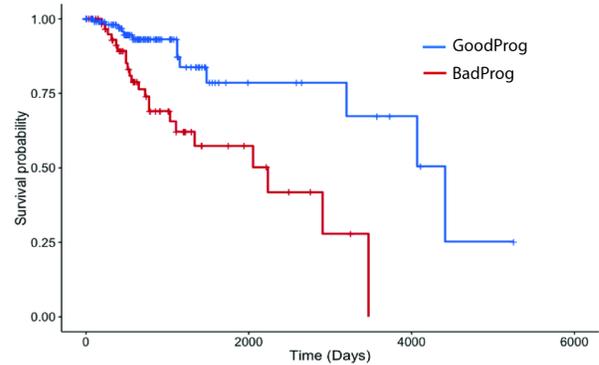
Fig 5. Expression of a subset of lncRNAs is associated with survival in LGG patients.
(A) Schematic of survival analysis: 60% of LGG patients were randomly selected to be the test set used to find survival-associated lncRNAs using Cox regression analysis. A summed Cox coefficient derived from 64 survival-associated lncRNAs (selected as in S3 Fig) was used to stratify patients in the test set into two survival groups. This same set of 64 lncRNAs was then used to derive the summed Cox coefficient in the validation set to separate them into two survival groups. (B) Survival plot shows the effective separation of patients from the test set into two distinct survival groups, good prognosis (GoodProg) and poor prognosis (BadProg), based on each patient's summed Cox coefficient of the 64 lncRNAs (hazard ratio [HR] = 2.168, 95% CI = 1.765–2.807, $p < 0.001$). (C) The summed Cox coefficient for the same 64 lncRNAs is capable of separating patients from the validation set into two groups with very distinct survival probabilities (HR = 1.921, 95% CI = 1.333–2.767, $p < 0.001$). LGG, lower grade glioma; lncRNA, long noncoding RNA.
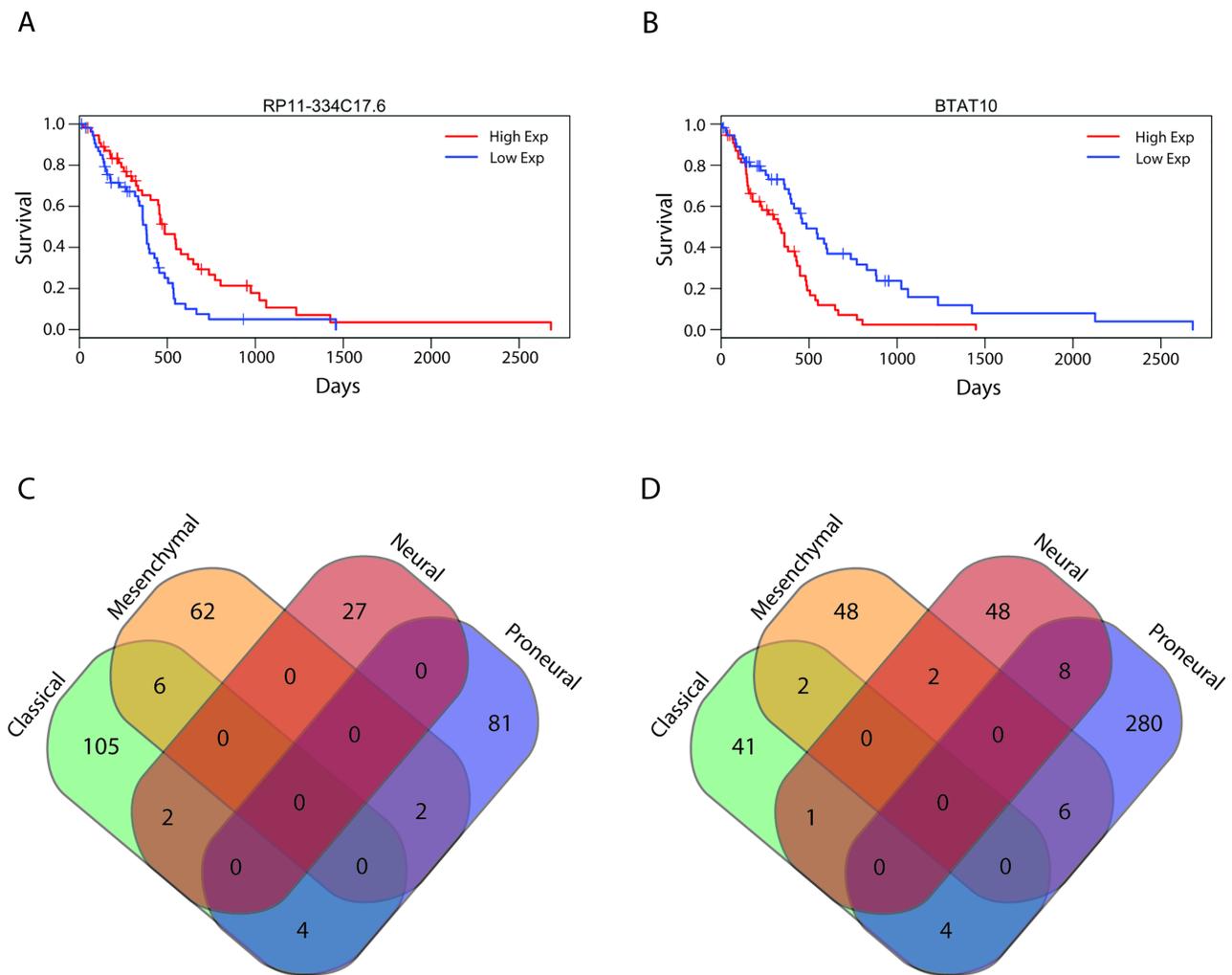
Fig 6. Survival-associated lncRNAs in GBMs and GBM subtypes.
(A and B) Representative survival plots of lncRNAs that predict survival in GBMs: RP11-334C17.6 (HR = 0.728, 95% CI = 0.6011–0.883, p = 0.00122) (A) and BTAT10 (HR = 1.298, 95% CI = 1.0881–1.548, p = 0.00374) (B). (C and D) lncRNAs associated with a poor prognosis (C) and good prognosis (D) in individual subtypes show minimal overlap between subtypes. GBM, glioblastoma multiforme; HR, hazard ratio; lncRNA, long noncoding RNA.

Good Prognosis Group		Poor Prognosis Group	
Pathways	Number of lncRNAs	Pathways	Number of lncRNAs
Intracellular signaling cascade	67	Response to external stimulus	59
Protein kinase cascade	61	Chromosome organization and biogenesis	55
Response to external stimulus	60	RNA processing	54
Organ development	58	Cell cycle (GO:0007049)	53
G protein coupled receptor protein signaling pathway	56	Cell cycle process	53
Cell surface receptor linked signal transduction	56	Cell cycle phase	52
Defense response	56	Defense response	52
Signal transduction	56	Mitotic cell cycle	51
Protein amino acid phosphorylation	55	Immune response	51
Regulation of signal transduction	55	RNA metabolic process	50

lncRNA, long noncoding RNA.

Table 1. Enriched biological pathways associated with prognostic lncRNAs by guilt-by-association analysis.

67
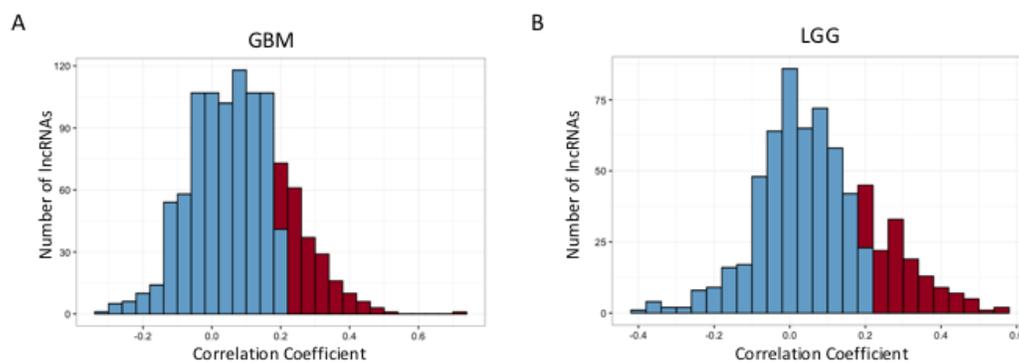
S1 Fig. Association of dysregulated lncRNA expression and tumor copy number variation.
(A) Histogram of Spearman correlation coefficients for lncRNAs and CNV in GBMs. (B) Histogram of Spearman correlation coefficients for lncRNAs and CNV in LGGs. Red lines indicate Spearman correlation coefficient greater than or equal to 0.2. Blue lines indicate non-correlated lncRNAs.
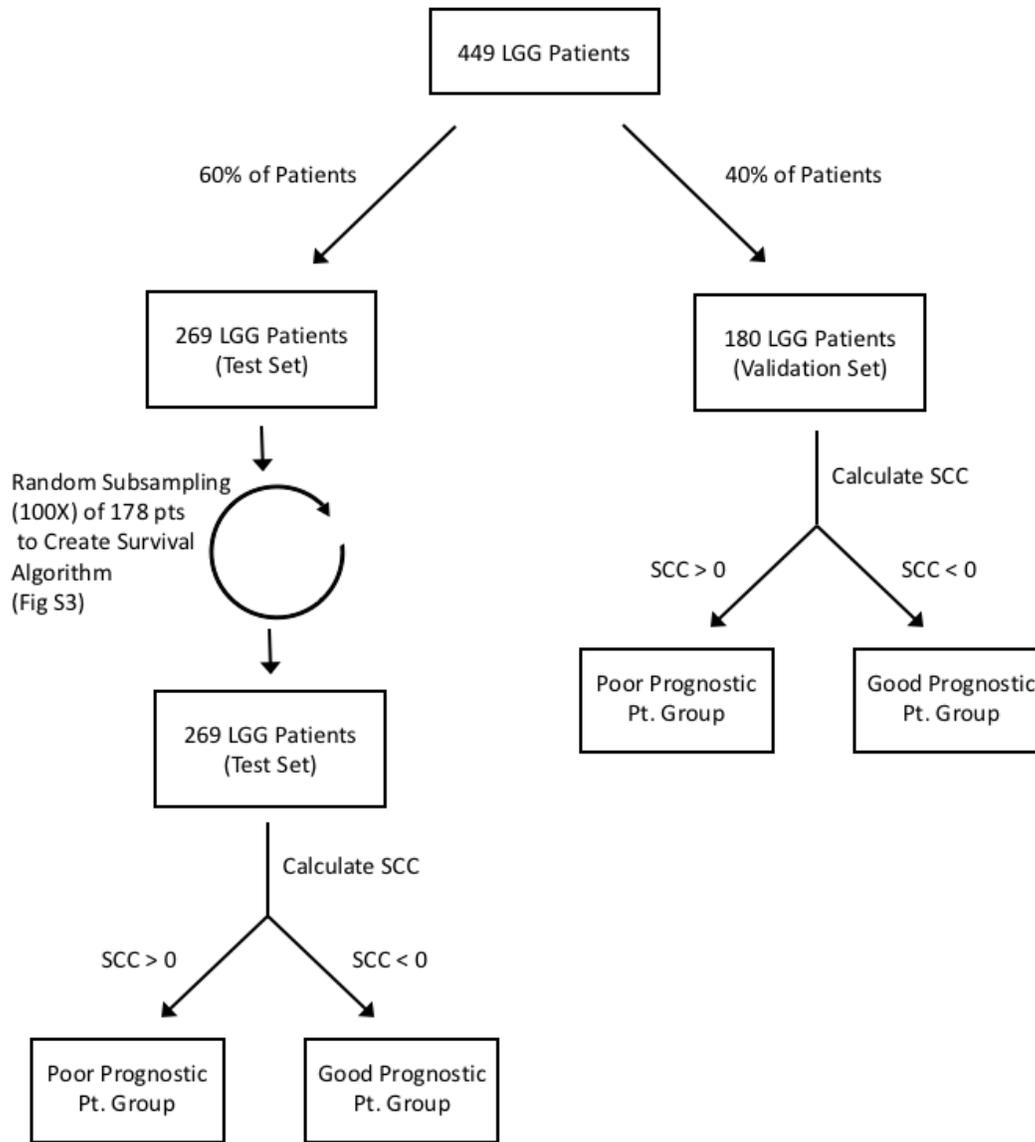
S2 Fig. Schematic of patient separation for survival algorithm development and validation.
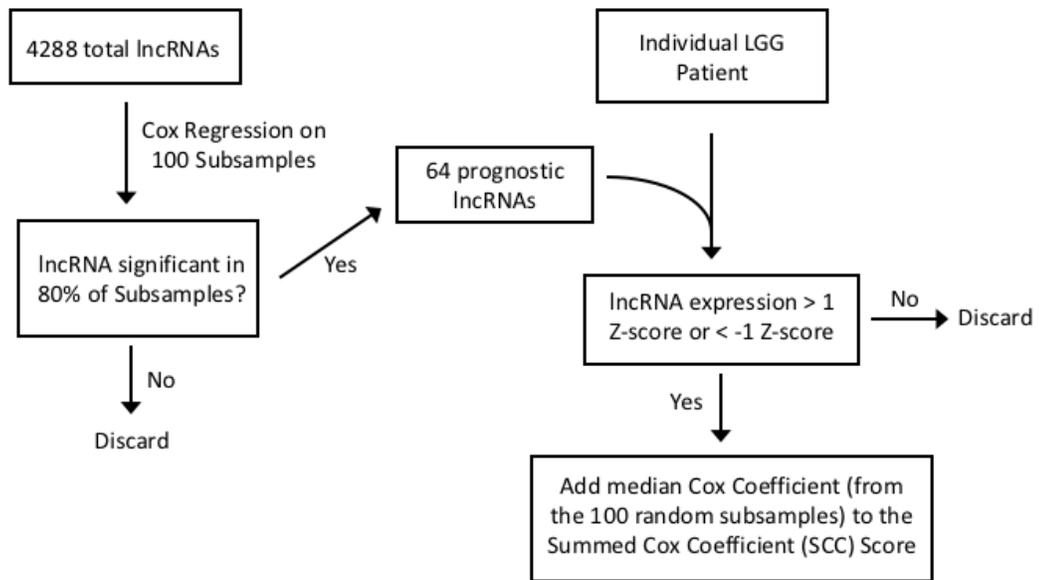
S3 Fig. Schematic for creating survival algorithm using lncRNA expression and Cox regression.

Chapter 3

LINC00152 Promotes Invasion in Glioblastomas and is Upregulated and Associated with Poor Outcome in Diverse Tumor Types

Brian J Reon [1,2], Bruno Takao Real Karia [2] and Anindya Dutta [2]

[1]Department of Pathology, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA

[2]Department of Biochemistry, University of Virginia, Charlottesville, VA, USA

Abstract

Glioblastomas (GBMs) are the most common type of primary malignant brain tumor and is extremely aggressive, with median survival times of less than 16 months [1]. Long noncoding RNAs (lncRNAs) have increasingly been shown to contribute to oncogenesis in a wide variety of tumors, however their roles in brain cancers are not as well characterized. Here, we identify LINC00152 as being upregulated in GBMs and aggressive IDH1/2 wt grade II and III gliomas. High expression of LINC00152 is also associated with poor patient outcomes in both these tumor types. LINC00152 is a mainly cytoplasmic, evolutionarily conserved lncRNA, with a distant homolog present in mouse. Inhibition of LINC00152 expression results in a decrease in cellular invasion *in vitro* and overexpression of LINC00152 leads to a reciprocal increase in cellular invasion. RNA-seq revealed that genes that are increased following LINC00152 knockdown are enriched in ribosome and translation pathways. Secondary structure analysis of LINC00152 reveals a protein bound stem loop that contributes to LINC00152 invasive properties. Through

analyzing TCGA expression data, we have shown that LINC00152 is upregulated in over 10

cancer types and is associated with a poor prognosis in 7 tumor types. Taken together, this

suggests that LINC00152 is associated with aggressive tumors perhaps by promoting cellular

invasion.

Introduction

GBM (glioblastoma) are highly aggressive grade IV gliomas and are the most common type of malignant glioma, with 10,000 new diagnoses each year [2]. GBMs are a heterogeneous group of tumors that can be separated into four different subtypes, mesenchymal, classical, proneural and neural, based on their transcriptional profile. Most of the focus on understanding glioma tumor biology has been on studying protein coding genes and microRNAs [3]. These efforts have led to identifying commonly altered signaling pathways in GBMs, including mutations in EGFR, p53 and mTOR signaling [4,5]. Furthermore, microRNAs have been shown to play a role in many of the oncogenic phenotypes of GBMs, such as invasiveness and stemness of GBM stem cells [6,7]. Although there has been much effort on creating new targeted therapies for GBMs, focusing on some of the aforementioned pathways, most have not been effective and the standard of care therapy, a combination of maximal surgical resection, radiotherapy and Temozolomide, still leaves patients with an abysmal 5-year survival rate of roughly 10% [8].

High throughput sequencing efforts have revealed that a majority of the human genome, which was long thought to be transcriptionally silent is actually expressed. Indeed, when surveyed across many different cell types it was found that nearly 80% of the human genome is actually transcribed [9]. Many of these newly discovered transcripts are lncRNAs (long noncoding RNAs). LncRNAs are a class of ncRNAs that are longer than 200 bases in length and can be further subdivided into several subclasses based on genomic positioning relative to other genes as well as whether or not the lncRNA locus overlaps with an enhancer or other genomic regulatory element. One subclass of lncRNAs are lincRNAs (long intergenic noncoding RNAs) which reside in the intergenic genomic space and do not overlap with any other gene. LincRNAs have been shown to play many different functional roles in the cell, in part through regulation of

transcription, mRNA stability and mRNA translational efficiency [10–12]. Most of the research into the role of ncRNA in GBMs has been on microRNAs, with relatively few studies on lncRNAs. This leaves a crucial gap in our understanding of glioma pathogenesis. Indeed, lncRNAs have been shown to function in critical roles in a variety of tumor types, e.g. HOTAIR in breast cancer, SChLAP1 in aggressive prostate cancer,MALAT1 in lung cancer and DRAIC in seven different cancer types [12–15].

LINC00152 is a lincRNA that was first identified as being hypomethylated during hepatocellular carcinoma tumorigenesis [16]. Since its initial discovery in hepatocellular carcinoma, it has been shown to be dysregulated in gastric cancer and esophageal squamous cell carcinoma [17,18]. However, there are conflicting reports on exactly how LINC00152 functions to promote the invasive phenotype. One study has argued that LINC00152 directly interacts with EGFR and affects AKT signaling while others have suggested that LINC00152 acts as a ceRNA through titrating miR-18a-5p [19,20]. Recently, we identified LINC00152 through an in-depth genomic analysis of gliomas as being highly expressed in GBMs [21]. In this study we characterize LINC00152's association with GBM clinical features, tumor cell invasion as well as begin to functionally characterize LINC00152 structurally and identify potential protein interaction partners. Furthermore, we found that LINC00152 is overexpressed in 9 other tumor types compared to matched normal tissue and high LINC00152 expression is associated with a negative prognosis in 5 of these tumors.

Expression of LINC00152 in Aggressive Gliomas

We first identified LINC00152 from a comprehensive analysis of lncRNAs in gliomas [21]. LINC00152 was one of the most differentially expressed lncRNAs in GBMs compared to normal brain tissue, however it is not upregulated in grade II and III gliomas (Fig 3.1a and Sup

Fig 3.1a). We have validated the upregulation of LINC00152 in an independent set of patient

GBMs compared to normal FFPE brain tissue [21]. It should be noted that LINC00152 is a close

homolog of another lncRNA on chromosome 2, MIR4435-2HG, and both have nearly identical

sequences (98.8% sequence overlap) (Sup Fig 3.2a). It is therefore extremely difficult to discern

whether the expression differences we are detecting through RNA-seq and RT-PCR are a

reflection of LINC00152, MIR4435-2HG or both lncRNAs expression levels.

We tested whether LINC00152 is preferentially expressed in a particular GBM subtype,

but that did not appear to be the case. The differences in LINC00152 expression between the

subtypes were not statistically significant, although the median expression of LINC00152 is

lowest in the proneural GBM subtype ($p<0.1$) (Sup Fig 3.1b). Interestingly, LINC00152 is more

highly expressed in the IDHwt LGG subtype and even though LINC00152 is not upregulated in

LGGs as a whole, this LGG subtype expresses 4 times as much LINC00152 as normal brains (p

< 0.00001) (Fig 3.1b). This is interesting, because IDHwt LGGs are far more aggressive than the

other LGG subtypes and display clinical properties similar to GBMs [22]. We next wanted to

determine whether LINC00152 is conserved across species by looking at the syntenic region in

mouse, but we were not able to identify a mouse homolog in this region. However, we were able

to identify a mouse homolog, Gm14005, located in the MIR4435-2HG syntenic region that has

64.9% sequence identity to LINC00152, demonstrating that LINC00152 is evolutionarily

conserved (Fig 3.1c and Sup Fig 3.2b).

LINC00152 Expression Predicts Survival in GBMs and LGGs

Because LINC00152 is upregulated in brain tumors compared to normal brain tissue, we

next asked whether LINC00152 expression can be used to predict survival in GBM and LGG

patients. To do this, we analyzed RNA expression and clinical data from the TCGA for both

GBMs and LGGs. We then separated patients into two equal groups, the top 50% "High

Expressers" and the bottom 50% "Low Expressers", based on the expression levels of

LINC00152. In GBMs, patients who had high expression of LINC00152 had a worse prognosis

(p = 0.0039) compared to the Low Expressers group, with a median survival of 11.9 and 15.4

months, respectively (Fig 3.2a). Furthermore, LINC00152 expression was also able to separate

patients into two distinct prognostic groups in LGGs. LGG patients with high expression of

LINC00152 had a median survival of 62.1 months, while the low expressing group had a median

survival 98.2 months (p = 1.4 e-5) (Fig 2b). These results demonstrate that not only is

LINC00152 overexpressed in gliomas, but that this overexpression is associated with worse

patient outcome.

LINC00152 Expression Controls GBM cell Invasion

We next sought to determine whether the upregulation of LINC00152 seen in GBMs is

associated with any cancer phenotypes in GBM cell lines. LINC00152 has previously been

shown to affect multiple cellular phenotypes, including cell growth, migration, invasion and

epithelial-to-mesenchymal transition (EMT) [20,23,24].We knocked down LINC00152

expression using two separate siRNAs and found that LINC00152 knockdown did not affect cell

growth (Fig 3.3a). We next assayed whether LINC00152 expression was associated with tumor

cell invasion using a transwell migration assay. Knockdown of LINC00152 in U87 cell lines led

to a statistically significant reduction in cell invasion with both siRNAs targeting LINC00152

(Fig 3.3b). To further validate the role of LINC00152 in GBM cell invasion, we measured

invasion in U87 cells stably overexpressing LINC00152 and found that overexpression of

LINC00152 led to an increase of over 2 fold in the amount of invaded cells (Fig 3.3c). We also

found that LINC00152 is primarily localized in the cytoplasm of U87 cells (Sup Fig 3.3). These

findings suggest that LINC00152 upregulation in GBMs could in part be promoting the invasive phenotype that is commonly seen in patient tumors.

LINC00152 Expression Profile

In order to better understand how LINC00152 affects cellular invasion we performed RNA-seq on U87 following knocking down of LINC00152 using two different siRNAs. Knock down of LINC00152 lead to large changes in gene expression, with 184 RNAs being upregulated and 130 downregulated (Fig 3.4a). Examination of the differentially expressed gene list did not reveal any immediate insights into potential pathways that LINC00152 might affect. In order to get a better understanding of what pathways LINC00152 might affect, we performed GSEA (gene set enrichment analysis), a method that can identify pathway enrichment from RNA-seq and microarray [25,26]. Surprisingly, 9 of the top 15 significantly enriched gene sets were positively associated with the ribosome (e.g. genes upregulated by LINC00152 knock down were enriched in ribosome gene sets), suggesting that LINC00152 might be involved in regulating translation (Fig 3.4b and Sup Table 3.1).

Secondary Structure Components of LINC00152

Unlike proteins, which for the most part are composed of several well characterized domains, lncRNAs do not have well defined structural domains. This can complicate the interpretation of lncRNA structure function studies because a given deletion mutant could lead to misfolding of the RNA and subsequent misinterpretation of the functional significance of a particular region of the lncRNA. Over the past decade several new technologies have been developed to examine the secondary structures of lncRNAs on a global basis, one such technique is PARIS (psoralen analysis of RNA interactions and structures) [27]. PARIS is based on reversibly crosslinking RNA duplexes (stems of stem loops) and gentle digestion with a

ssRNase, S1 nuclease, to cut looped single stranded portions of an RNA's secondary structure. The surviving RNA duplexes from the stems are then ligated to each other and subjected to high throughput sequencing. RNAs containing stem-loops will have sequencing reads corresponding to the stems with gaps (corresponding to the loops) that do not overlap with a splice site. We analyzed publicly available PARIS data from HeLa and 293T cells to determine whether LINC00152 contains any secondary structure elements that could be detected by PARIS. Following alignment, we identified reads with a 2-nt gap that were present in both of the HeLa PARIS libraries, but not in any of the 293T samples (Sup Fig 3.4a). These reads are positioned from position 284 to 373 of the 496 nt long LINC00152, with a small 2 base gap starting at position 342. Sequence analysis of this region revealed a modest degree of complementarity, suggesting that this region might in fact form a stem loop structure (data not shown).

We next sought to get a better understanding of overall LINC00152 secondary structure. To do this, we used publicly available RNA secondary structure prediction tool, RNAstructure, to identify secondary structure predictions for LINC00152 that are consistent with a stem-loop being present from 284-373 [28]. The top 2 secondary structures with the lowest free energy differed in their exact base-pairing, but the overall stem loop structure was largely the same. Importantly, both structures were consistent with a stem loop being present from position 284 to 373 (Fig 3.5a, Sup Fig 3.5a & b). Furthermore, the resulting loop from the stem formation is rather small, 4 nt, which is consistent with the small 2 nt gap seen by PARIS.

We next asked if we could use a separate method to independently validate the hairpin formation in LINC00152. Ribo-seq (Ribosome profiling) is a technique that has been used to identify RNAs that interact with the ribosome and how the ribosome is distributed across those RNAs [29]. This information has in part been used to ascertain that while some lncRNAs are

associated with ribosomes, the ribosome distribution is not consistent with a pattern of translating ribosomes[30]. Recently it was determined that the experimental preparation of isolating polysomes for Ribo-seq does not remove all other RNA-protein complexes and often have traces of other RBPs (RNA binding proteins) and their bound RNAs. This resulting RNA footprints from the RBPs can be detected in Ribo-seq and can be used to determine sites of non-ribosome bound proteins to the RNAs [31]. To determine if we could identify RBP-RNA footprints from LINC00152 we analyzed publicly available Ribo-seq data from normal brain samples [32]. In two out of the three normal brain Ribo-seq samples we detected an RBP footprint at positons 300-335 of LINC00152. In addition, in one of the samples there was an RBP footprint from 354-380 (Fig 3.5a and Sup Fig 3.4b). These two peaks are located at the same stem loop that was detected by PARIS, and peaks from both techniques overlap almost entirely (Fig 3.5a).

LINC00152 Structure Function Studies

In order to determine whether the newly identified, potentially protein bound, stem loop plays a role in LINC00152 function, we created two deletion constructs. The first removed the minimal amount of the protein bound stem loop, nts 280-401 (M2), and another that removed the stem loop and the remaining 3' end (M3) (Fig 3.5a). The sites of the deletions were chosen based on PARIS and Ribo-seq analysis as well as identifying unbase-paired regions of LINC00152 that were consistent between two *in silico* predicted structures of LINC00152 (Fig 3.5b and Sup Fig 3.4a & b). We next assessed whether overexpression of the mutants was able to effect U87 cell invasion. Overexpression of M2 led to an increase in cell invasion, however the increase was less than what was seen with full-length LINC00152 ($p < 0.05$) (Fig 3.5c). In addition, overexpression of M3 also led to an increase in cell invasion but to a lower extent then full-

length LINC00152 and M2 (Fig 3.5c). These results suggest that the stem-loop plays a partial role in the invasive properties of LINC00152, but there are other elements in the 5' end of LINC00152 that also contribute to its invasive functions.

LINC00152 in other Cancers

In addition to playing a role in GBM invasion, other studies examining the function of LINC00152 have found that it is involved in cell growth and invasion in other tumor types as well [23,24]. Furthermore, other groups have shown that LINC00152 is overexpressed in gastric cancer and hepatocellular carcinoma [17,23]. We next wanted to ask how frequently is LINC00152 upregulated in cancers compared to their respective normal tissues. To do this, we compared the expression of LINC00152 in all TCGA tumor samples with paired normal and tumor RNA-seq data. Surprisingly, LINC00152 is upregulated in nearly every tumor type we analyzed, including breast invasive carcinoma, colorectal carcinoma, head and neck squamous cell carcinoma, renal clear cell and papillary, hepatocellular carcinoma, stomach adenocarcinoma, uterine carcinoma, lung adenocarcinoma and thyroid carcinoma (Fig 3.6a-j). Prostate adenocarcinoma was the only tumor type that had statistically significant decrease in the levels of LINC00152 in the tumor samples compared to normal (data not shown).

Since LINC00152 is overexpressed in the majority of tumors that we have analyzed, we next wanted to determine whether LINC00152 expression is associated with patient survival in the TCGA tumors that had higher levels of LINC00152 compared to the paired normal samples. To do this, we performed Kaplan Meier analysis for each tumor type by separating patients into two groups, the top quartile LINC00152 expressing tumors and the lowest quartile LINC00152 expressing tumors. From the original list of tumors, LINC00152 expression was associated with poor patient outcome in head and neck squamous cell carcinoma, renal clear cell carcinoma,

hepatocellular carcinoma and lung adenocarcinoma (Fig 3.7 a-d). The poorer outcome of patients

with renal papillary carcinoma was not statistically significant comparing the top and bottom

quartiles of LINC00152 expression (p = 0.1), but the poorer outcome was statistically significant

(p = 0.014) when we compared patients in the top third and bottom third based on LINC00152

expression (Sup Fig 3.6a).

Although LINC00152 was not overexpressed in LGGs relative to normal brain, it was

upregulated in an aggressive subpopulation of LGGs (those with IDH wild type) and was

associated with poor patient outcome. This made us realize that even if a tumor type does not

overexpress LINC00152 globally relative to normal tissue, overexpression of the lncRNA in

specific tumors may still be associated with poor outcome. We therefore examined other TCGA

tumors which did not show a global increase of LINC00152 expression in the cancers relative to

normal tissue for the predictive value of the expression of this lncRNA. Interestingly, even

among these tumors, LINC00152 expression was associated with poor patient outcome in

pancreatic adenocarcinoma and acute myeloid leukemia, when we comparing the tumors in the

top quartile and bottom quartile for LINC00152 expression (Sup Fig 3.6b and c). These results

highlight the fact that in nine tumor types (GBMs, LGGs, head and neck squamous cell

carcinoma, renal clear cell carcinoma, hepatocellular carcinoma, lung adenocarcinoma, renal

papillary carcinoma, pancreatic adenocarcinoma and acute myeloid leukemia) LINC00152

appears to function as an oncogene whose expression is associated with a poor patient outcome.

Discussion

The human genome was once thought to be mainly dormant and that most of the

transcription was devoted to producing protein coding genes. We now know that the genome is

transcriptionally vibrant and only a small fraction of the expressed genome, roughly 2%, encodes

for protein coding genes. GWAS and high throughput sequencing studies have found that many of the genomic lesions and expression alterations seen in cancer and other pathologies fall within non-protein coding regions of the genome and may lead to dysregulation of ncRNAs [33–36]. Furthermore, there is a growing body of evidence implicating lncRNAs in playing a direct role in normal cellular physiology, as well as driving pathogenesis in a variety of disorders, including cancer [37–41]. Indeed, recent work has illustrated the critical role that lncRNAs play in cancer, including iconic examples such as HOTAIR in breast cancer and HULC in hepatocellular carcinoma and DRAIC in seven different cancer types [12,15,42].

In this study we have shown that LINC00152 is a lincRNA that is upregulated in many different cancer types and is highly upregulated in GBMs. Although LINC00152 is not upregulated in all LGGs relative to normal brain tissue, it is upregulated in the highly malignant IDHwt LGG subtype, further supporting LINC00152's association with aggressive tumors. This raised the interesting possibility that in tumors where LINC00152 is not differentially over-expressed or is moderately upregulated in the tumor population relative to normal tissue, LINC00152 could still be highly upregulated in a more aggressive subgroup of the tumors. This was indeed found to be true in Pancreatic Adenocarcinomas and Acute Myeloid Leukemias. LINC00152 expression is associated with patient survival in nine different cancer types, including GBMs and LGGs. LINC00152 expression promotes cell invasion, which is consistent with its association with poor patient outcomes.

LINC00152 is an evolutionarily conserved lincRNA that appears to be the product of a gene duplication from MIR4435-2HG, as the latter has synteny with the mouse homolog, Gm14005. Although we cannot rule out that the phenotypes that we observe with siRNA directed towards LINC00152 are acting in part through knocking down the highly similar

MIR4435-2HG, we do not believe that affecting MIR4435-2HG levels fully explains LINC00152 effect on invasion. Firstly, we see an increase in cell invasion when we exogenously express LINC00152, arguing that it is an inherent effect of LINC00152 and not MIR4435-2HG. Secondly, although LINC00152 is nearly 99% identical to MIR4435-2HG, the mouse homolog shares more sequence homology with LINC00152 than MIR4435-2HG, even though it lies in the syntenic region of the human genome (data not shown). Also, note that the reads aligning to LINC00152 in RNA-seq and Ribo-seq experiments matched with LINC00152 even at the positions not shared with MIR4435-2HG, arguing that LINC00152 is expressed in the tumors and cell lines and is being directly measured in the PARIS and Ribo-seq datasets. Interestingly, GSEA of RNA-seq from LINC00152 knocked down cells revealed that LINC00152 might be involved in translation. Although, RNA-seq of LINC00152 overexpressed cells is needed to better understand and validate this connection.

In conclusion, we have identified LINC00152 as being strongly upregulated in GBMs and aggressive LGGs as well as 10 other cancers. Furthermore, we have shown that LINC00152 expression is associated with survival in many of these tumors. In GBM cell lines, LINC00152 is involved in promoting cellular invasion, in part through a protein-bound stem loop in its 3' end. Our data supports the notion that LINC00152 functions as an oncogenic lincRNA in GBMs and potentially plays a critical oncogenic role in a wide variety of cancer types. LINC00152 could also serve as a tumor biomarker or a target for future cancer therapeutics.

Materials and Methods

Expression of LINC00152 in TCGA datasets and Survival Analysis

The expression of LINC00152 in GBMs and LGGs compared to normal brain and tumor subtypes was performed as previously described [21]. Expression of LINC00152 in all

83

other TCGA tumors was determined by comparing expression data of only those tumors that had a matched normal tissue sample. Statistical significance was determined using a paired t-test. TCGA patient survival data for GBMs and LGGs were retrieved from cBioPortal (www.cbioportal.org) and survival data from the remaining tumor types were retrieved from OncoLnc (www.oncolnc.org) on 12/2016 [43–45]. The expression threshold used to separate patients are outlined in the main text. Kaplan Meier plots and p-values were generated using the 'survminer' package for R.

Cell growth and Invasion Analysis

U87 cells were treated with two different siRNA against LINC00152 (siL152-2: 5'-UGACACACUUGAUCGAAUA-3', siL152-3: 5'-CCGGAAUGCAGCUGAAAGA-3') or a nonspecific siGL2 control siRNA during two sequential rounds of transfection. Each 20 uM of siRNA was first reverse transfected into U87 with 2 ul of RNAimax transfection reagent (ThermoFisher). 24 hours later, we performed a second round of transfections using the same quantities of reagents except doing conventional transfections. 24 hours after the final transfection, cells were harvested and used for subsequent analysis. For transiently overexpressing LINC00152 or LINC00152 mutants in U87 cells, 1 ug of plasmid was transfected into U87 cells with 2 ul of Lipofectamine 2000 reagent (ThermoFisher). Cells were harvested after 48 hours for downstream analysis. For measuring cell growth, 2,000 cells were plated in quadruplets in 96 well plates and cell growth was measured using standard MTT reagent (Promega). To measure invasion, 10^5 U87 cells in serum free media were seeded into 24-well Matrigel Invasion Chambers (BD Biosciences) and the bottom was filled with media and 10% FBS as the chemoattractant. Cells were allowed to invade for 8 hours and then fixed and stained

with crystal violet/methanol and invaded cells were counted. LINC00152 subcellular localization was determined using Protein and RNA Isolation System (ThermoFisher).

RNA-seq Analysis

U87 cells were treated with siRNA as mentioned earlier and total cell RNA was isolated using TRIzol and subsequently purified using RNeasy Isolation kit (Qiagen). Sequencing libraries were generated using NEB NEXT Ultra directional RNA Library prep kit and samplers were barcoded with NEBNext Multiplexing oligos per standard manufacturer protocols. Libraries were sequenced with 50 bp paired-end reads on an Illumina HiSeq 2500. Sequencing reads were aligned to the hg38 reference genome using HISAT2 [46]. The two LINC00152 siRNA treated datasets were treated as replicates and gene abundances and identification of differentially expressed genes were performed using and DESeq2 [47,48]. GSEA analysis was performed on preranked gene list based on fold change (siLINC00152/siGL2) with 1000 permutations against the GO gene sets [26].

LINC00152 Structure Predictions

Secondary structure predictions of LINC00152 were determined using RNAstructure [28]. The two structures with the lowest predicted free energies were selected for comparisons with PARIS and Ribo-seq. For PARIS data analysis of LINC00152, raw sequencing data from Lu *et. al.* was aligned to the hg19 genome using STAR (spliced transcripts alignment to a reference) with the alignment parameters outlines in Lu *et. al.* [27,49]. Aligned reads were then processed to identify gapped mapping to LINC00152 and visualized with IGV [50]. We used ribosome profiling data from Gonzalez *et. al.* and aligned reads to the hg19 genome using HISAT2 [32]. We then examined reads that mapped to LINC00152 for their distribution along the message to ensure that they were not legitimate ribosome footprints using IGV [50]. The

predicted secondary structure elements and protein bound region were then compared to the *in silico* secondary structure predictions.

References

1.   Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. N Engl J Med [Internet]. 2005 Mar 10;352(10):987–96. Available from: http://dx.doi.org/10.1056/NEJMoa043330

2.   Ostrom QT, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, et al. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. Neuro-Oncology [Internet]. 2015 Oct 1;17(suppl 4):iv1-iv62. Available from: http://neuro-oncology.oxfordjournals.org/content/17/suppl_4/iv1.short

3.   Huse JT, Holland EC. Targeting brain cancer: advances in the molecular pathology of malignant glioma and medulloblastoma. Nat Rev Cancer [Internet]. 2010 May;10(5):319–31. Available from: http://dx.doi.org/10.1038/nrc2818

4.   Network CGAR. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008;455:1061–8.

5.   Brennan CW, Verhaak RGW, McKenna A, Campos B, Noushmehr H, Salama SR, et al. The Somatic Genomic Landscape of Glioblastoma. Cell. 2013 Oct;155(2):462–77.

6.   Møller HG, Rasmussen AP, Andersen HH, Johnsen KB, Henriksen M, Duroux M. A Systematic Review of MicroRNA in Glioblastoma Multiforme: Micro-modulators in the Mesenchymal Mode of Migration and Invasion. Mol Neurobiol [Internet]. 2013;47(1):131–44. Available from: http://dx.doi.org/10.1007/s12035-012-8349-7

7.   Deng Y, Zhu G, Luo H, Zhao S. MicroRNA-203 As a Stemness Inhibitor of Glioblastoma Stem Cells. Mol Cells [Internet]. 2016 Aug 31;39(8):619–24. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4990754/

8.   Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJB, Janzer RC, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol [Internet]. 2009 May;10(5):459–66. Available from: http://www.sciencedirect.com/science/article/pii/S1470204509700257

9.   Consortium TEP. An Integrated Encyclopedia of DNA Elements in the Human Genome. Nature [Internet]. 2012 Sep 6;489(7414):57–74. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3439153/

10.   Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature [Internet]. 2012;491(7424):454–7. Available from:

http://dx.doi.org/10.1038/nature11508

11.   Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K, et al. Control of somatic
       tissue differentiation by the long non-coding RNA TINCR. Nature. 2013
       Jan;493(7431):231–5.

12.   Gupta R a, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding
       RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature
       [Internet]. 2010;464(7291):1071–6. Available from:
       http://dx.doi.org/10.1038/nature08975

13.   Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, et al. The long noncoding
       RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF
       complex. Nat Genet [Internet]. 2013 Nov;45(11):1392–8. Available from:
       http://dx.doi.org/10.1038/ng.2771

14.   Gutschner T, Hämmerle M, Eißmann M, Hsu J, Kim Y, Hung G, et al. The non-coding
       RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells.
       Cancer Res [Internet]. 2013 Feb 1;73(3):1180–9. Available from:
       http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3589741/

15.   Sakurai K, Reon BJ, Anaya J, Dutta a. The lncRNA DRAIC/PCAT29 Locus Constitutes a
       Tumor-Suppressive Nexus. Mol Cancer Res [Internet]. 2015;13(5):828–38. Available
       from: http://mcr.aacrjournals.org/cgi/doi/10.1158/1541-7786.MCR-15-0016-T

16.   Neumann O, Kesselmeier M, Geffers R, Pellegrino R, Radlwimmer B, Hoffmann K, et al.
       Methylome analysis and integrative profiling of human HCCs identify novel
       protumorigenic factors. Hepatology [Internet]. 2012 Nov 1;56(5):1817–27. Available
       from: http://dx.doi.org/10.1002/hep.25870

17.   Cao W-J, Wu H-L, He B-S, Zhang Y-S, Zhang Z-Y. Analysis of long non-coding RNA
       expression profiles in gastric cancer. World J Gastroenterol [Internet]. 2013 Jun
       21;19(23):3658–64. Available from:
       http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3691033/

18.   Yang S, Ning Q, Zhang G, Sun H, Wang Z, Li Y. Construction of differential mRNA-
       lncRNA crosstalk networks based on ceRNA hypothesis uncover key roles of lncRNAs
       implicated in esophageal squamous cell carcinoma. Oncotarget; Adv Online Publ
       [Internet]. 2016;10. Available from:
       http://www.impactjournals.com/oncotarget/index.php?journal=oncotarget&amp

19.   Xia T, Liao Q, Jiang X, Shao Y, Xiao B, Xi Y, et al. Long noncoding RNA associated-
       competing endogenous RNAs in gastric cancer. Sci Rep [Internet]. 2014 Aug 15;4:6088.
       Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4133709/

20.   Zhou J, Zhi X, Wang L, Wang W, Li Z, Tang J, et al. Linc00152 promotes proliferation in
       gastric cancer through the EGFR-dependent pathway. J Exp Clin Cancer Res [Internet].
       2015 Nov 4;34:135. Available from:

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4632266/

21.   Reon BJ, Anaya J, Zhang Y, Mandell J, Purow B, Abounader R, et al. Expression of
       lncRNAs in Low-Grade Gliomas and Glioblastoma Multiforme: An In Silico Analysis.
       PLOS Med [Internet]. 2016 Dec 6;13(12):e1002192. Available from:
       http://dx.doi.org/10.1371%2Fjournal.pmed.1002192

22.   Network CGAR. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade
       Gliomas. N Engl J Med [Internet]. 2015 Jun 10;372(26):2481–98. Available from:
       http://dx.doi.org/10.1056/NEJMoa1402121

23.   Ji J, Tang J, Deng L, Xie Y, Jiang R, Li G, et al. LINC00152 promotes proliferation in
       hepatocellular carcinoma by targeting EpCAM via the mTOR signaling pathway.
       Oncotarget [Internet]. 2015 Dec 15;6(40):42813–24. Available from:
       http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4767473/

24.   Zhao J, Liu Y, Zhang W, Zhou Z, Wu J, Cui P, et al. Long non-coding RNA Linc00152 is
       involved in cell cycle arrest, apoptosis, epithelial to mesenchymal transition, cell
       migration and invasion in gastric cancer. Cell Cycle [Internet]. 2015 Aug 3;14(19):3112–
       23. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4825539/

25.   Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-
       1[alpha]-responsive genes involved in oxidative phosphorylation are coordinately
       downregulated in human diabetes. Nat Genet [Internet]. 2003 Jul;34(3):267–73. Available
       from: http://dx.doi.org/10.1038/ng1180

26.   Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene
       set enrichment analysis: A knowledge-based approach for interpreting genome-wide
       expression profiles. Proc Natl Acad Sci [Internet]. 2005 Oct 25;102(43):15545–50.
       Available from: http://www.pnas.org/content/102/43/15545.abstract

27.   Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, et al. RNA Duplex Map in
       Living Cells Reveals Higher-Order Transcriptome Structure. Cell [Internet]. 2016 May
       19;165(5):1267–79. Available from:
       http://www.sciencedirect.com/science/article/pii/S0092867416304226

28.   Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction
       and analysis. BMC Bioinformatics [Internet]. 2010;11(1):129. Available from:
       http://dx.doi.org/10.1186/1471-2105-11-129

29.   Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-Wide Analysis in
       Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science (80-)
       [Internet]. 2009 Apr 10;324(5924):218 LP-223. Available from:
       http://science.sciencemag.org/content/324/5924/218.abstract

30.   Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. Ribosome Profiling
       Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. Cell [Internet].
       2013 Jul 3;154(1):240–51. Available from:

http://www.sciencedirect.com/science/article/pii/S0092867413007113

31.   Ji Z, Song R, Huang H, Regev A, Struhl K. Transcriptome-scale RNase-footprinting of RNA-protein complexes. Nat Biotech [Internet]. 2016 Apr;34(4):410–3. Available from: http://dx.doi.org/10.1038/nbt.3441

32.   Gonzalez C, Sims JS, Hornstein N, Mela A, Garcia F, Lei L, et al. Ribosome Profiling Reveals a Cell-Type-Specific Translational Landscape in Brain Tumors. J Neurosci [Internet]. 2014 Aug 13;34(33):10924–36. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4131009/

33.   Mirza AH, Kaur S, Brorsson CA, Pociot F. Effects of GWAS-Associated Genetic Variants on lncRNAs within IBD and T1D Candidate Loci. PLoS One [Internet]. 2014 Aug 21;9(8):e105723. Available from: http://dx.doi.org/10.1371%2Fjournal.pone.0105723

34.   Kerin T, Ramanathan A, Rivas K, Grepo N, Coetzee GA, Campbell DB. A Noncoding RNA Antisense to Moesin at 5p14.1 in Autism. Sci Transl Med. 2012 Apr;4(128):128ra40-128ra40.

35.   Huarte M. The emerging role of lncRNAs in cancer. Nat Med [Internet]. 2015 Nov;21(11):1253–61. Available from: http://dx.doi.org/10.1038/nm.3981

36.   Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet [Internet]. 2015 Mar;47(3):199–208. Available from: http://dx.doi.org/10.1038/ng.3192

37.   Grote P, Wittler L, Währisch S, Hendrix D, Beisaw A, Macura K, et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev Cell [Internet]. 2013 Jan 28;24(2):206–14. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4149175/

38.   Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, et al. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of [beta]-secretase. Nat Med [Internet]. 2008 Jul;14(7):723–30. Available from: http://dx.doi.org/10.1038/nm1784

39.   Garzon R, Volinia S, Papaioannou D, Nicolet D, Kohlschmidt J, Yan PS, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc Natl Acad Sci USA [Internet]. 2014 Dec 30;111(52):18679–84. Available from: http://www.pnas.org/content/111/52/18679.abstract

40.   Flockhart RJ, Webster DE, Qu K, Mascarenhas N, Kovalski J, Kretz M, et al. BRAFV600E remodels the melanocyte transcriptome and induces BANCR to regulate melanoma cell migration. Genome Res [Internet]. 2012 Jun 1;22(6):1006–14. Available from: http://genome.cshlp.org/content/22/6/1006.abstract

41.   Van Grembergen O, Bizet M, de Bony EJ, Calonne E, Putmans P, Brohée S, et al. Portraying breast cancers with long noncoding RNAs. Sci Adv [Internet]. 2016 Sep

2;2(9):e1600220. Available from:
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5010371/

42.   Panzitt K, Tschernatsch MMO, Guelly C, Moustafa T, Stradner M, Strohmaier HM, et al.
        Characterization of HULC, a Novel Gene With Striking Up-Regulation in Hepatocellular
        Carcinoma, as Noncoding RNA. Gastroenterology [Internet]. 2007 Jan;132(1):330–42.
        Available from: http://www.sciencedirect.com/science/article/pii/S0016508506017859

43.   Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer
        Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics
        Data. Cancer Discov [Internet]. 2012 May 9;2(5):401 LP-404. Available from:
        http://cancerdiscovery.aacrjournals.org/content/2/5/401.abstract

44.   Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative
        Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci
        Signal [Internet]. 2013 Apr 2;6(269):pl1 LP-pl1. Available from:
        http://stke.sciencemag.org/content/6/269/pl1.abstract

45.   Anaya J. OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs.
        PeerJ Comput Sci. 2016;2(e67).

46.   Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory
        requirements. Nat Meth [Internet]. 2015 Apr;12(4):357–60. Available from:
        http://dx.doi.org/10.1038/nmeth.3317

47.   Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput
        sequencing data. Bioinformatics [Internet]. 2014 Sep 25;31(2):166–9. Available from:
        http://dx.doi.org/10.1093/bioinformatics/btu638

48.   Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
        RNA-seq data with DESeq2. Genome Biol [Internet]. 2014;15(12):550. Available from:
        http://dx.doi.org/10.1186/s13059-014-0550-8

49.   Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
        universal RNA-seq aligner. Bioinformatics [Internet]. 2013 Jan 25;29(1):15–21. Available
        from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3530905/

50.   Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al.
        Integrative genomics viewer. Nat Biotech [Internet]. 2011 Jan;29(1):24–6. Available
        from: http://dx.doi.org/10.1038/nbt.1754
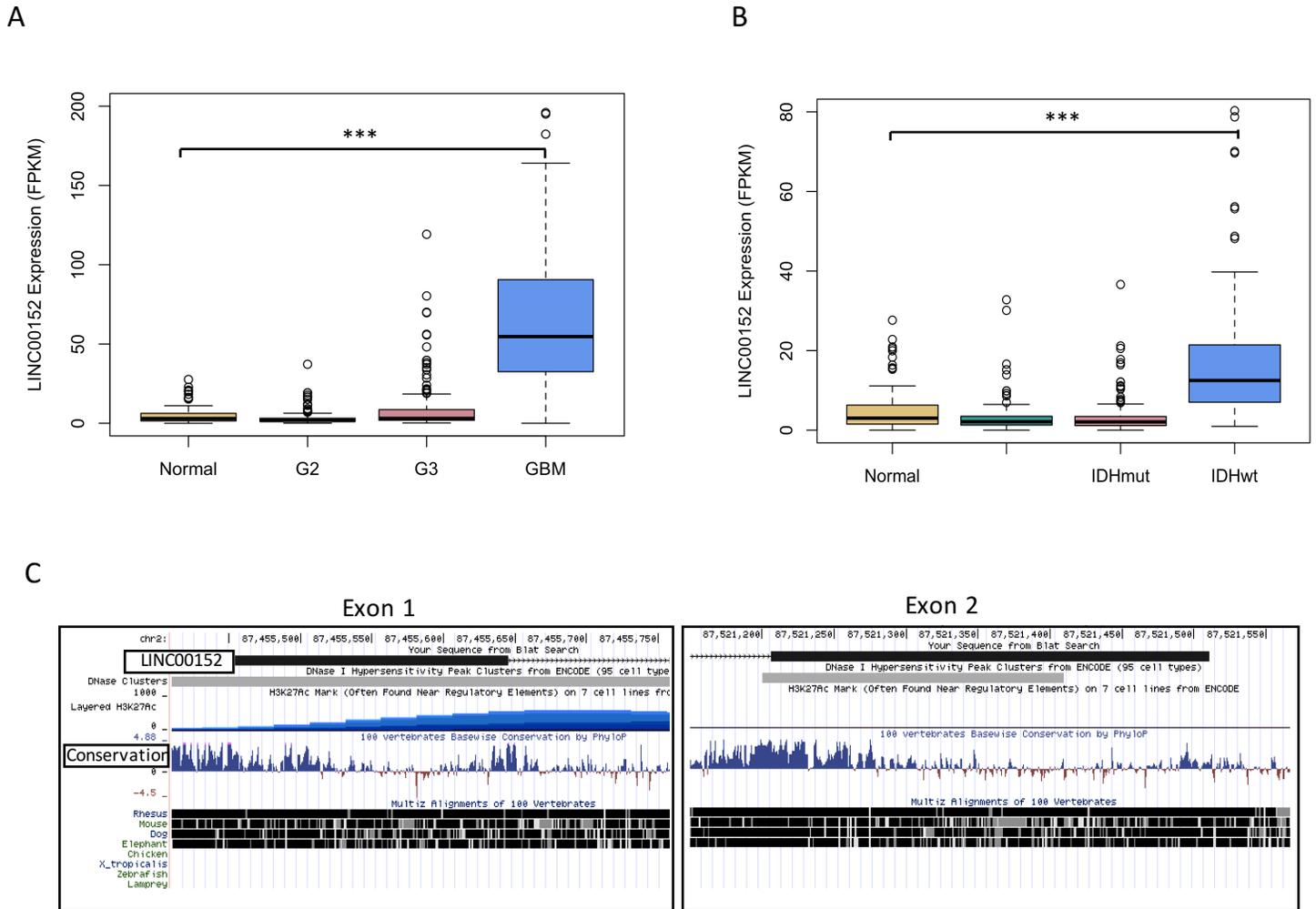
Figure 3.1. LINC00152 is an evolutionarily conserved lncRNA upregulated in aggressive gliomas. A) Boxplot of LINC00152 expression in GBM, G3 (grade 3 glioma), G2 (grade 2 glioma) and normal brain tissue. B) Boxplot of LINC00152 expression in LGG subtypes and normal brain tissue. C) Snap shot of exon 1 and 2 of LINC00152 from the UCSC genome browser with the vertebrate conservation track highlighted. *** p-value < 0.0001

A

B
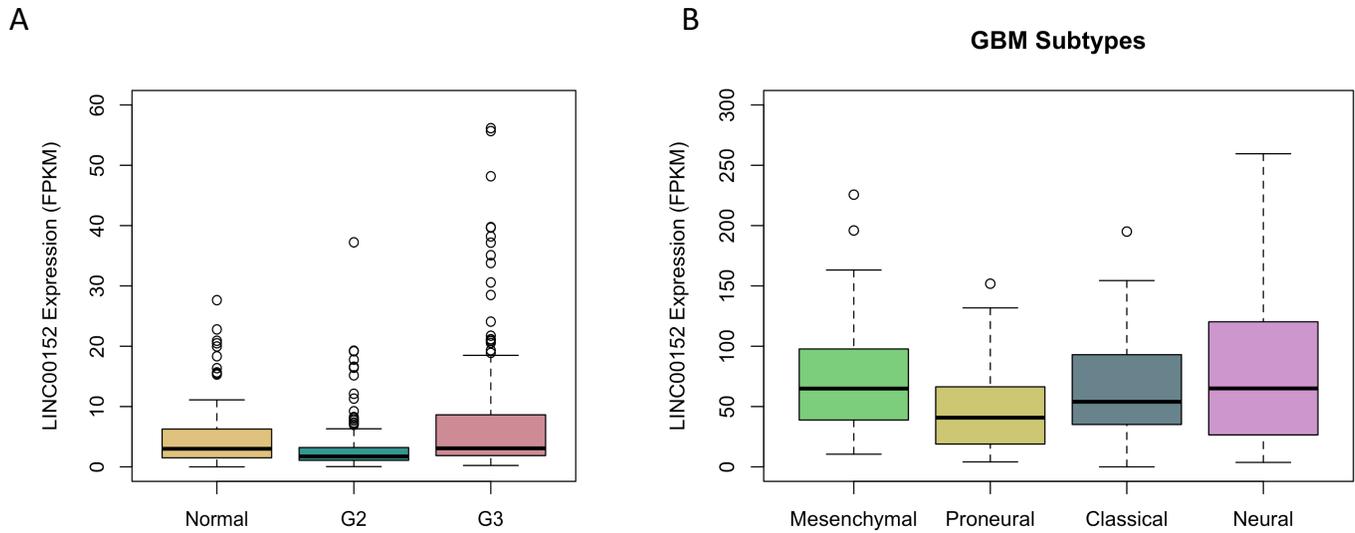
GBM Subtypes

Supplemental Figure 3.1. LINC00152 is not differentially expressed in LGGs globally or between GBM subtypes. A) Boxplot of LINC00152 expression in G3 (grade 3 glioma), G2 (grade 2 glioma) and normal brain tissue. B) Boxplot of LINC00152 expression in GBM subtypes.

A

```
LINC00152      1 GTGCGCCTTTTTTTTTTTTTTCCTTCTTAGTCGTGTGTACATCATTGGGAA    50
                 ||||||| ||||||||||||||||||||||||||||||||||||||||||
MIR4435-2HG   13 GTGCGCC-TTTTTTTTTTTTTCCTTCTTAGTCGTGTGTACATCATTGGGAA   61

LINC00152     51 TGGAGGGAAATAAATGACTGGATGGTCGCTGCTTTTTAAGTTTCAAATTG   100
                 |||||||||||||||||||||||||||||||||||||||||||||||||
MIR4435-2HG   62 TGGAGGGAAATAAATGACTGGATGGTCGCTGCTTTTTAAGTTTCAAATTG   111

LINC00152    101 ACATTCCAGACAAGCGGTGCCTGAGCCCGTGCCTGTCTTCAGATCTTCAC   150
                 |||||||||||||||||||||||||||||||||||||||||||||||||
MIR4435-2HG  112 ACATTCCAGACAAGCGGTGCCTGAGCCCGTGCCTGTCTTCAGATCTTCAC   161

LINC00152    151 AGCACAGTTCCTGGGAAGGTGGAGCCACCAGCCTCTCCTTGAATAACTGG   200
                 ||||||||||||||||||||||||||||||||||||||| |||||||||
MIR4435-2HG  162 AGCACAGTTCCTGGGAAGGTGGAGCCACCAGCCTCTCCCTGAATAACTGG   211

LINC00152    201 GAGATGAAACAGGAAGCTCTATGACACACTTGATCGAATATGACAGACAC   250
                 |||||||||||||||||||||||||||||||||||||||||||||||||
MIR4435-2HG  212 GAGATGAAACAGGAAGCTCTATGACACACTTGATCGAATATGACAGACAC   261

LINC00152    251 CGAAAATCACGACTCAGCCCCCTCCAGCACCTCTACCTGTTGCCCGCCGA   300
                 .||||||||||||||| |||||||||||||||||||||||||||||||||
MIR4435-2HG  262 TGAAAATCACGACTCATCCCCCTCCAGCACCTCTACCTGTTGCCCGCCGA   311

LINC00152    301 TCACAGCCGGAATGCAGCTGAAAGATTCCCTGGGGCCTGGTTCCAACCGC   350
                 ||||||||||||||||||||||||||||||||||||||||||||||.||
MIR4435-2HG  312 TCACAGCCGGAATGCAGCTGAAAGATTCCCTGGGGCCTGGTTCCAACTGC   361

LINC00152    351 CCACTGTGGACTCTGAGGCCTCTGCATTTGCGGGTGGTCTGCCTGTGATA   400
                 |||||||||||||||||||||||||||||||||||||||||||||||||
MIR4435-2HG  362 CCACTGTGGACTCTGAGGCCTCTGCATTTGCGGGTGGTCTGCCTGTGATA   411

LINC00152    401 TTTTGGTCATGGGCTGGTCTGGTCGGTTTCCCATTTGTCTGGCCAGTCTC   450
                 |||||||||||||||||||||||||||||||||||||||||||||||||
MIR4435-2HG  412 TTTTGGTCATGGGCTGGTCTGGTCGGTTTCCCATTTGTCTGGCCAGTCTC   461

LINC00152    451 TATGTGTCTTAATCCCTTGTCCTTCATTAAAAGCAAAACTAAAGAA      496
                 .|||||||||||||||||||||||||||||||||||||||||||||
MIR4435-2HG  462 TGTGTGTCTTAATCCCTTGTCCTTCATTAAAAGCAAAACTAAAGAA      507
```

B

```
LINC00152      5 GCCTTTTTTTTTTTTTTCCTTCTTAGTCGTGTGTACATCATTGGGAATGGA    54
                 .|||||||||||||||   |||.||.||.|||||.|||||..||.|||| ||
GM14005       79 gtctttttttttttttt--ttcctaatcatgtggacatctctgagaat-ga   125

LINC00152     55 GGGAAATAAATGACTGGATG--------GTCGCTGCTTTTTAAGTTTCAA    96
                 |||        |||||||.||        ||.-||||||||.|||||||||
GM14005      126 ggg--------gactgggtgactttgcagtgtctgctttctaagtttcaa   167

LINC00152     97 ATTGACATTCCAGACAAGCGGTGCCTGAGCCCGTGCCTGTCTTCAGATCT   146
                 |.||||.||||||| ||||.|  |||          ||.|.||||..|||||
GM14005      168 agtgactttccaggcaagc--tgc----------gcatatcttgggatct   205

LINC00152    147 TCACAGCACAGTTCCTGGGA----AGGTGGAGCC----ACCAGCCTCTCC   188
                 |||||||||||||||||||||    |||.||||-||    .|||||||||||
GM14005      206 tcacagcacagtggctgggaaggcaggcggacccgtctatcagcttctcc   255

LINC00152    189 TTGAATAACTGGGAGATGAAACAGGAAGCTCTATGACACACTTGATCGAA   238
                 .|||||||||||||.|||||||||||||.|||||.|||||||||||||||
GM14005      256 ctgaataactgggagctgaaacaggaagccccatgacacagttgatcgaa   305

LINC00152    239 TATG--ACAGACACCG-AAAATCACGAC-TCAGCCCCCTCCAGCACCTCT   284
                 .|||  |||||.|||| |||||||||||  |||||||||||||||||||||
GM14005      306 aatgacacagacacagaaaaatcaggacgt--gccccagcagcacccct   353

LINC00152    285 ACCTGTTGCCCGCCGATCACAGCCGGAATGCAGCTGAAAGATTCCCTGGG   334
                 |||||.||||.||||.||.||||||||||||||||||||||    ||.|||    |
GM14005      354 acctgctgcctgctgatcacggctgg-------------gaggccc---g   387

LINC00152    335 GCCTGG--TTCCAACCGCCCACTGTGGACTCTGAGGC--CTC--TGCATT   378
                 .||.|| |.|||.|||...|   |||||||.|.|| ||| |||  |
GM14005      388 tccagggttcccacccgggaac---ggactctta-gctactcagtgc--t   431

LINC00152    379 TGCGGGTGGTCTGCCTG-TGATATTTT-GGTCATG---------GGCTGG   417
                 ..||||.||||.||.|||| .||.||||| .||.|||           .|||||
GM14005      432 cacgggtgctgggcctgcagagcttttcgg--atgctctcttacagctgg   479

LINC00152    418 TCTGGTCGGTTT-----------CCCATTTGTC---------------TG   441
                 |.||.|...|||       |..||||.|| |
GM14005      480 tttgttatcttttaatcaatcaactaattaatctattactttatttattg   529

LINC00152    442 GCCAGTCTCTATGTGTCTTAATCCCTTGTCCTTCATTAAAAGCAAAACTA   491
                 |||   |||....|||||||||.||||||||||||||||||||||||||..|
GM14005      530 gcc---ctccccatgtcttgatcctttt---cttcattaaaagcaaaagga   573

LINC00152    492 AAGAA      496
                 |||||
GM14005      574 aagaa      578
```

Supplemental Figure 3.2. LINC00152 is highly similar to the lncRNA MIR4435-2HG and has an evolutionarily conserved homolog in mouse. A) Sequence alignment of LINC00152 and MIR4435-2HG. B) Sequence alignment of LINC00152 and mouse homolog Gm14005

A   GBM

B   LGG
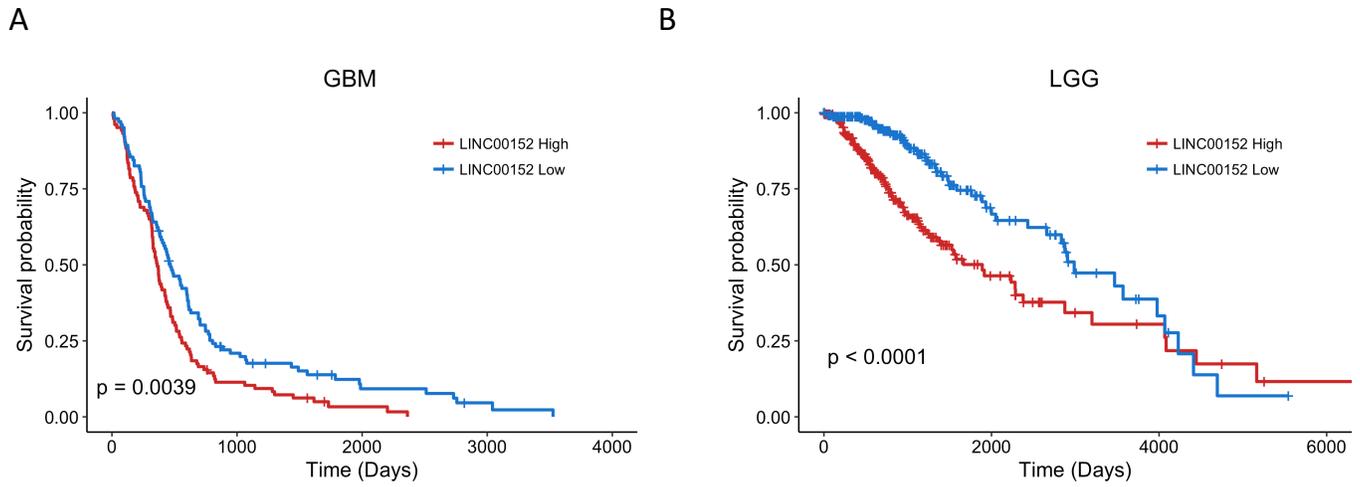
Figure 3.2. LINC00152 expression is associated with patient prognosis in GBMs and LGGs.
A) Kaplan Meier of GBM patients separated into the 50% highest expressing LINC00152
cohort and the lower 50% expressing cohort. B) Kaplan Meier of LGG patients separated into
the 50% highest expressing LINC00152 cohort and the lower 50% expressing cohort.
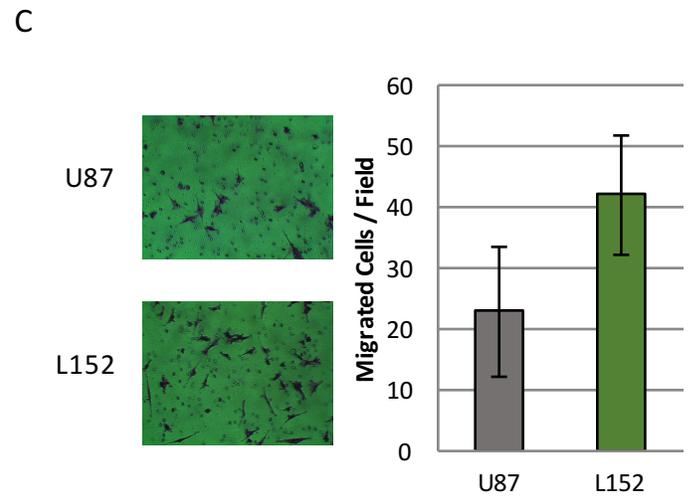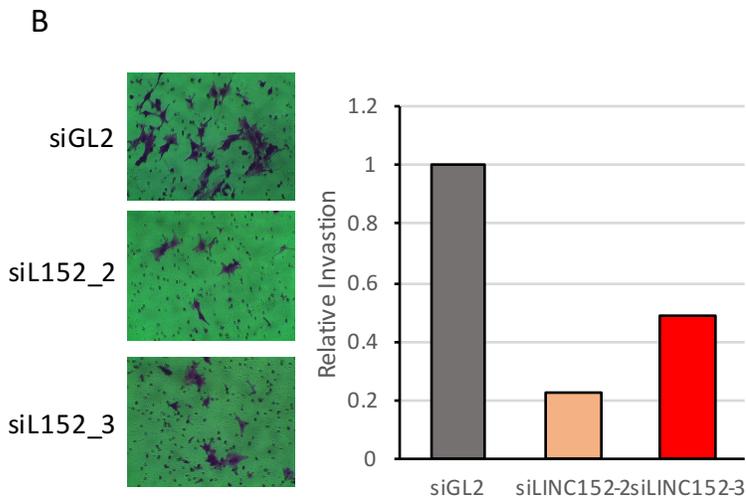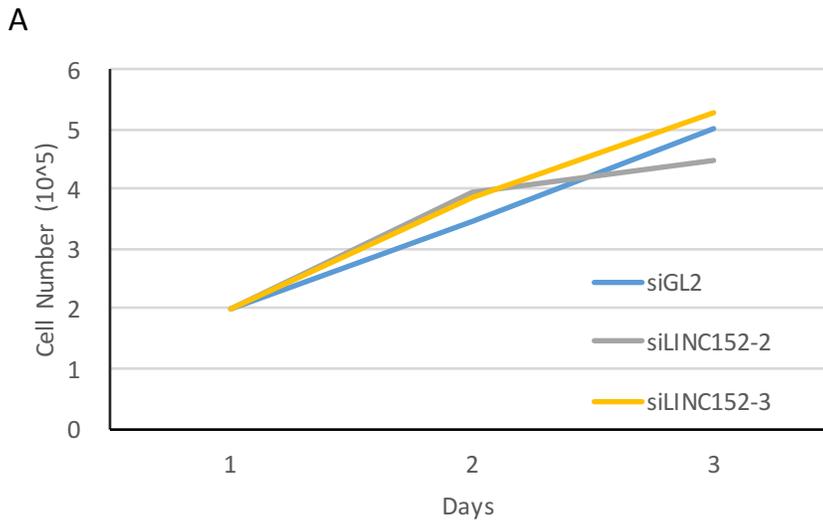
Figure 3.3. LINC00152 expression promotes cell invasion. A) Cell growth analysis of U87 cells treated with siRNA against LINC00152. B) U87 cell invasion after treatment with siRNA against LINC00152. C) U87 cell invasion with U87 cells stably overexpressing LINC00152

A

U87

N   C

Lamin

a-Tubulin

B
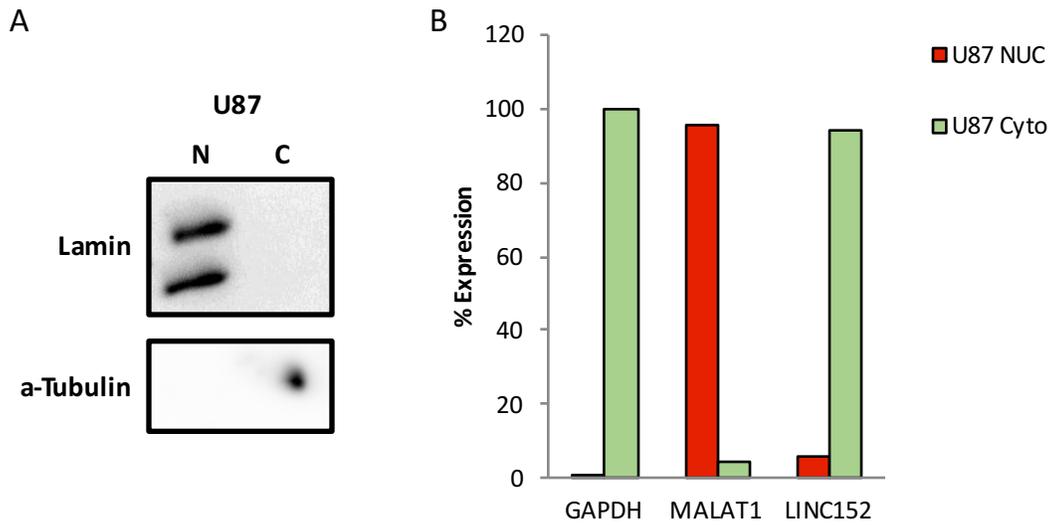
Supplemental Figure 3.3. LINC00152 is primarily localized in the cytoplasm. A) Western blot of Lamin and a-Tubulin, markers of the nucleus and cytoplasm, respectively. B) RT-PCR of LINC00152 and a cytoplasmic RNA marker, GAPDH, and a nuclear RNA marker, MALAT1.
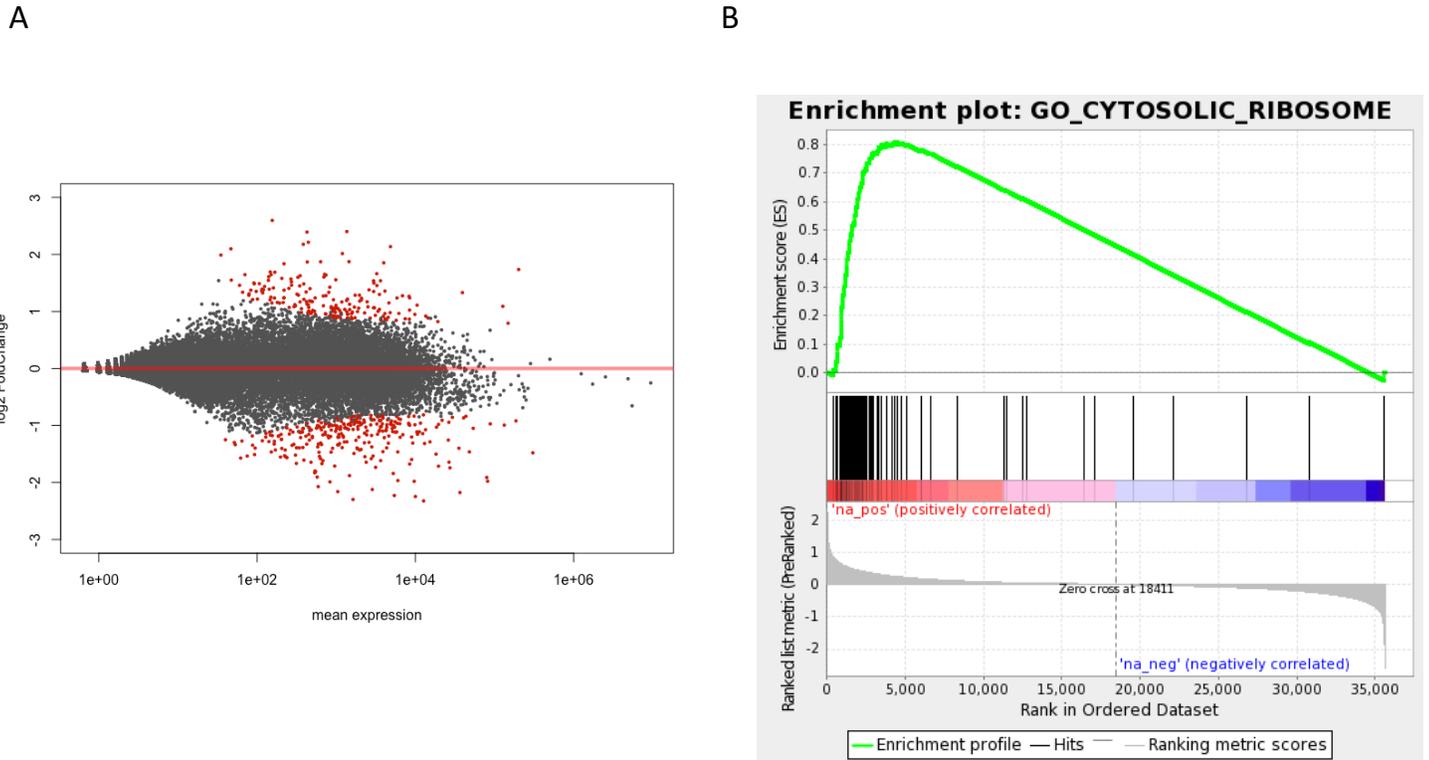
Figure 3.4. LINC00152 regulates many genes that are involved in translation. A) MA plot of differentially expressed genes following knockdown of LINC00152. Red dot are genes that were differentially expressed (FDR < 0.05). B) Representative enrichment plot from GSEA analysis of LINC00152 changed genes.

Gene Ontology Terms	NES	FDR q-value
GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM	2.84	> 0.001
GO_CYTOSOLIC_RIBOSOME	2.83	> 0.001
GO_RIBOSOMAL_SUBUNIT	2.8	> 0.001
GO_TRANSLATIONAL_INITIATION	2.77	> 0.001
GO_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM	2.67	> 0.001
GO_PROTEIN_TARGETING_TO_MEMBRANE	2.65	> 0.001
GO_CYTOSOLIC_LARGE_RIBOSOMAL_SUBUNIT	2.62	> 0.001
GO_MULTI_ORGANISM_METABOLIC_PROCESS	2.59	> 0.001
GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY	2.58	> 0.001
GO_RIBOSOME	2.56	> 0.001
GO_STRUCTURAL_CONSTITUENT_OF_RIBOSOME	2.56	> 0.001
GO_CYTOSOLIC_SMALL_RIBOSOMAL_SUBUNIT	2.56	> 0.001
GO_LARGE_RIBOSOMAL_SUBUNIT	2.55	> 0.001
GO_SMALL_RIBOSOMAL_SUBUNIT	2.54	> 0.001
GO_RRNA_METABOLIC_PROCESS	2.46	> 0.001

Supplemental Table 3.1. Table of the top 15 most positively enriched gene sets following LINC00152 knockdown
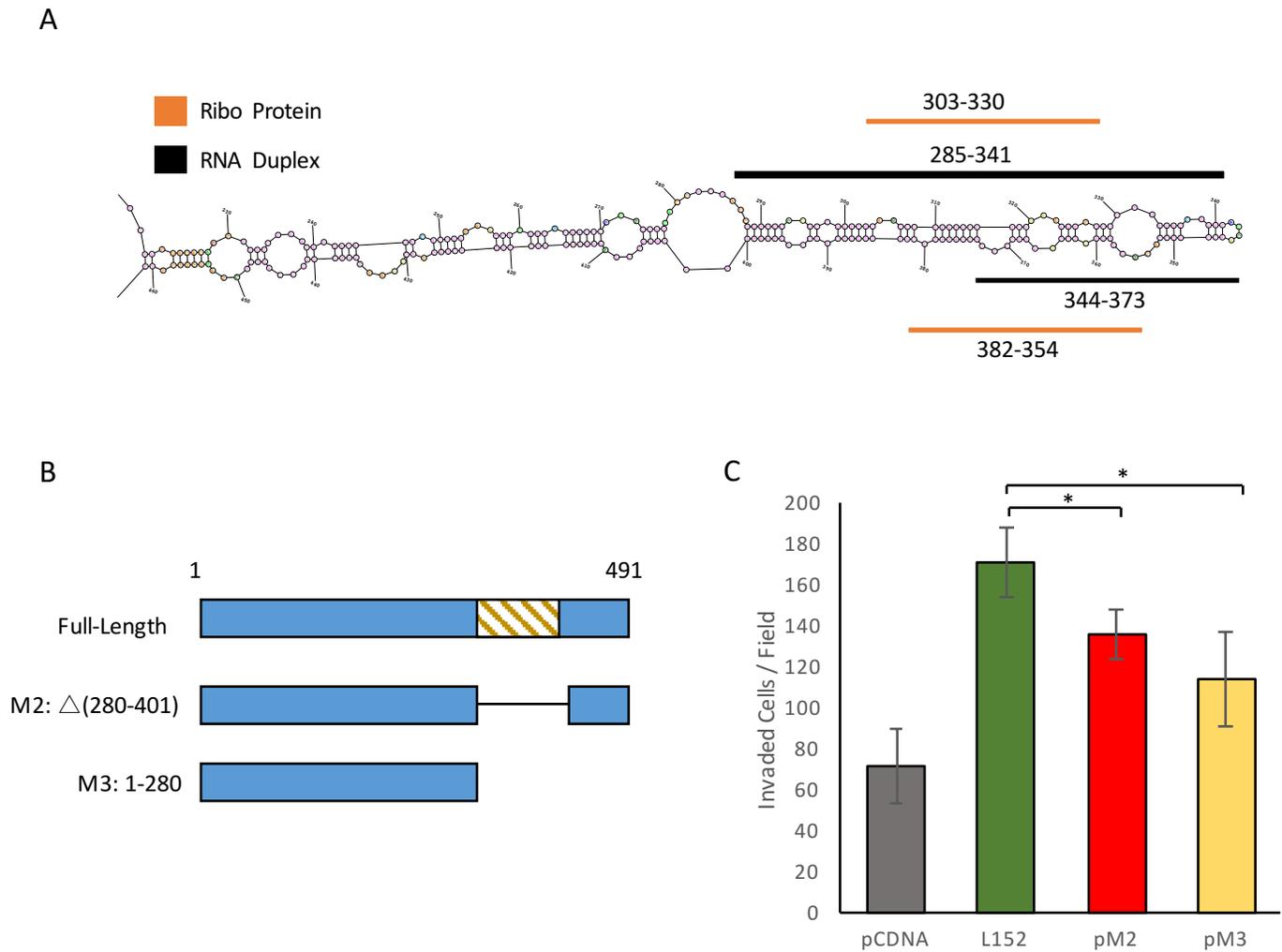
Figure 3.5. A protein-bound stem loop in LINC00152 is involved in U87 invasion. A) Predicted secondary structure of LINC00152 and the stem loop and protein bound regions identified by PARIS and Ribo-seq, respectively. B) LINC00152 mutant maps. C) U87 cell invasion after overexpression of LINC00152 or the deletion mutants.

A

LINC00152 Exon 2

B
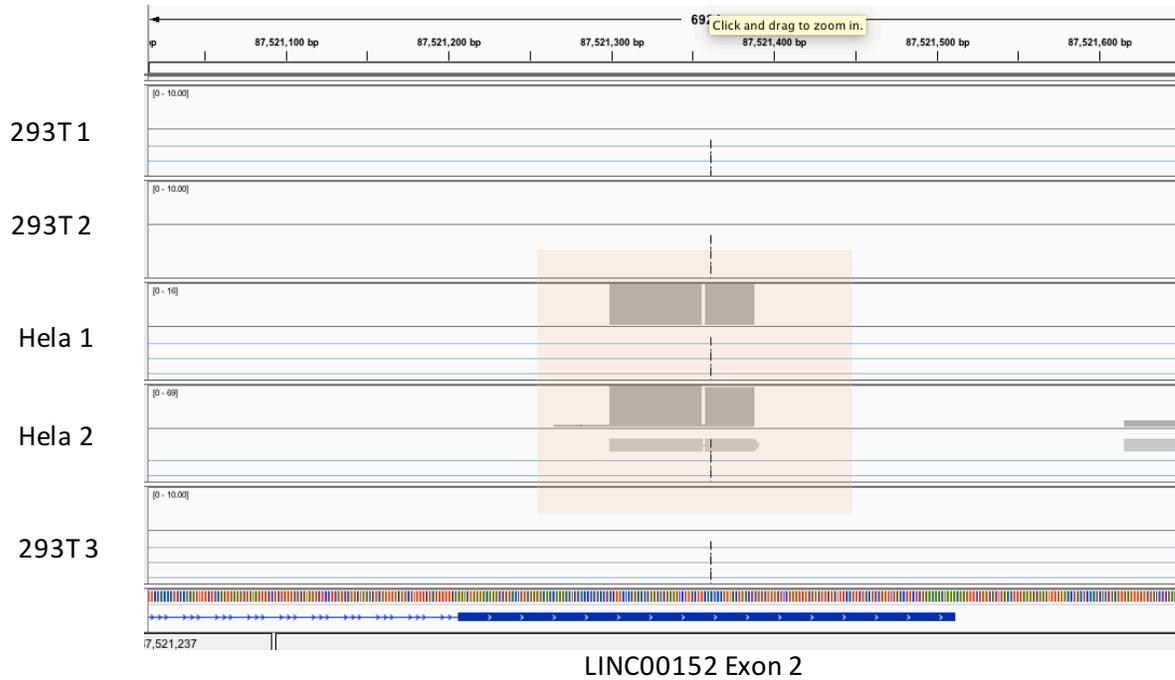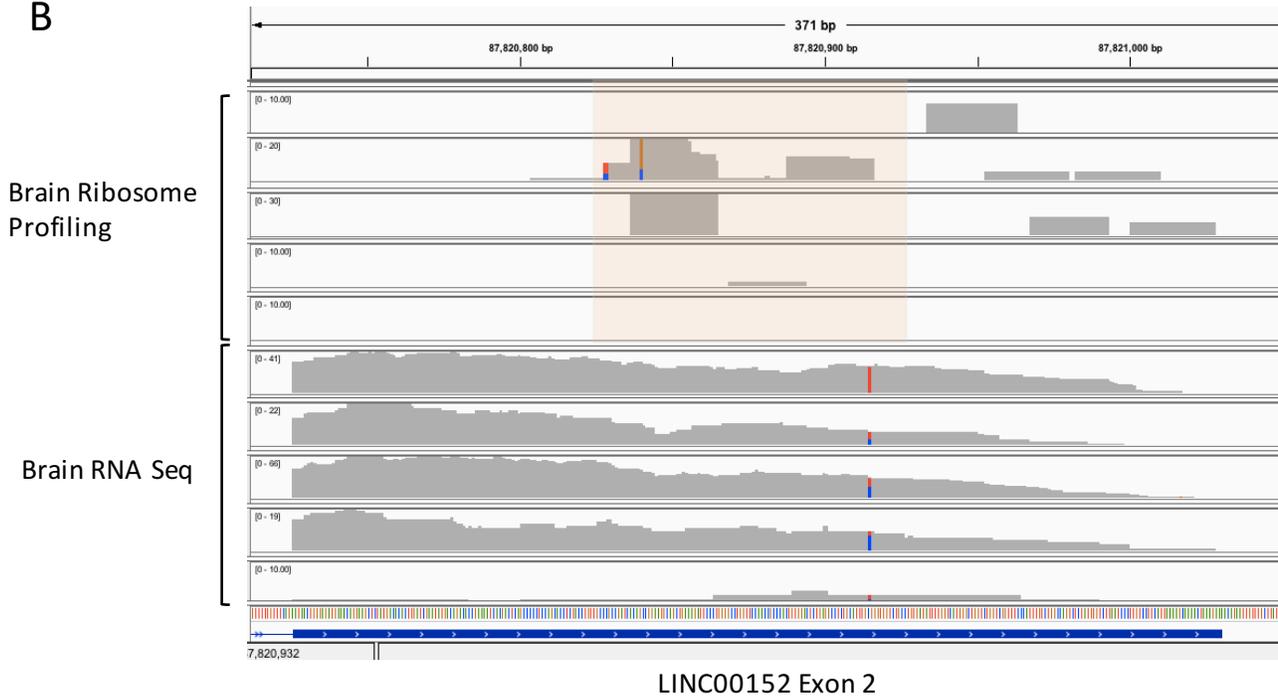
LINC00152 Exon 2

Supplemental Figure 3.4. IGV view of reads mapping to LINC00152 from PARIS and Ribo-seq A) A 2-bp gapped read is present in both HeLa PARIS libraries that maps to exon 2 of LINC00152. B) IGV view of RNA-seq and Ribo-seq reads that map to exon 2 of LINC00152. Beige shaded areas show regions of overlap between A and B.
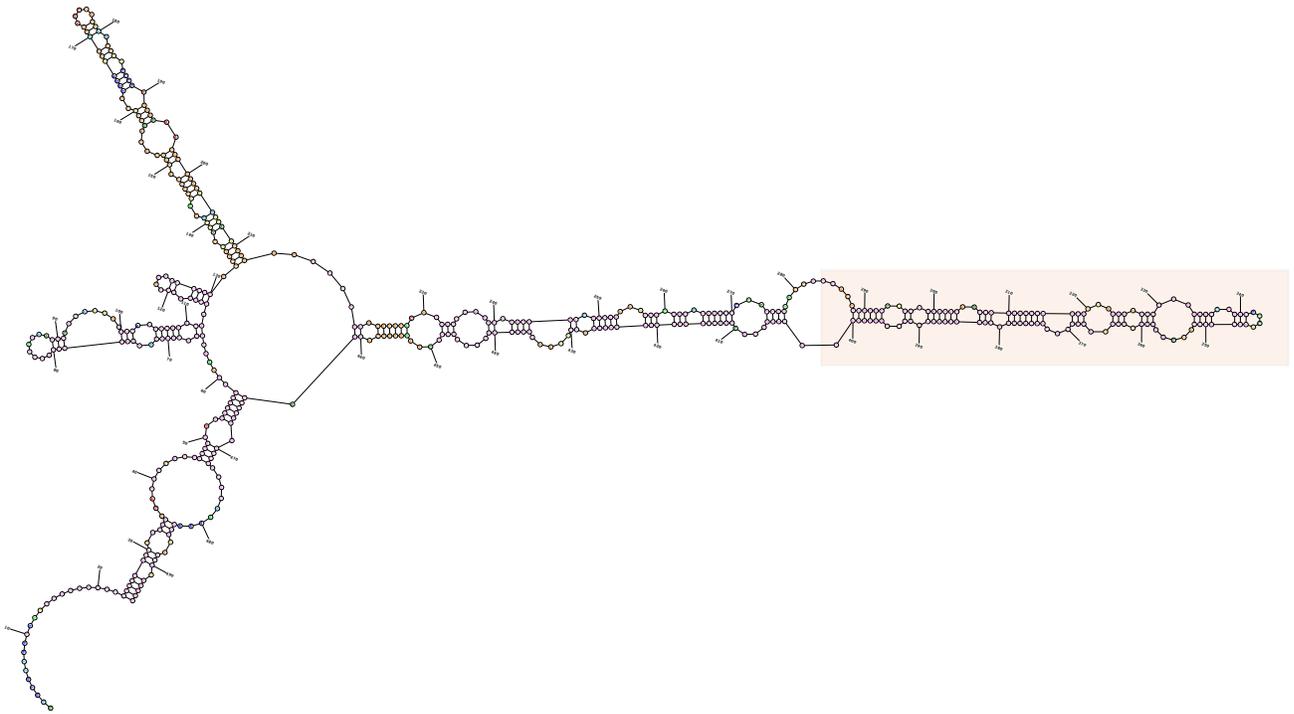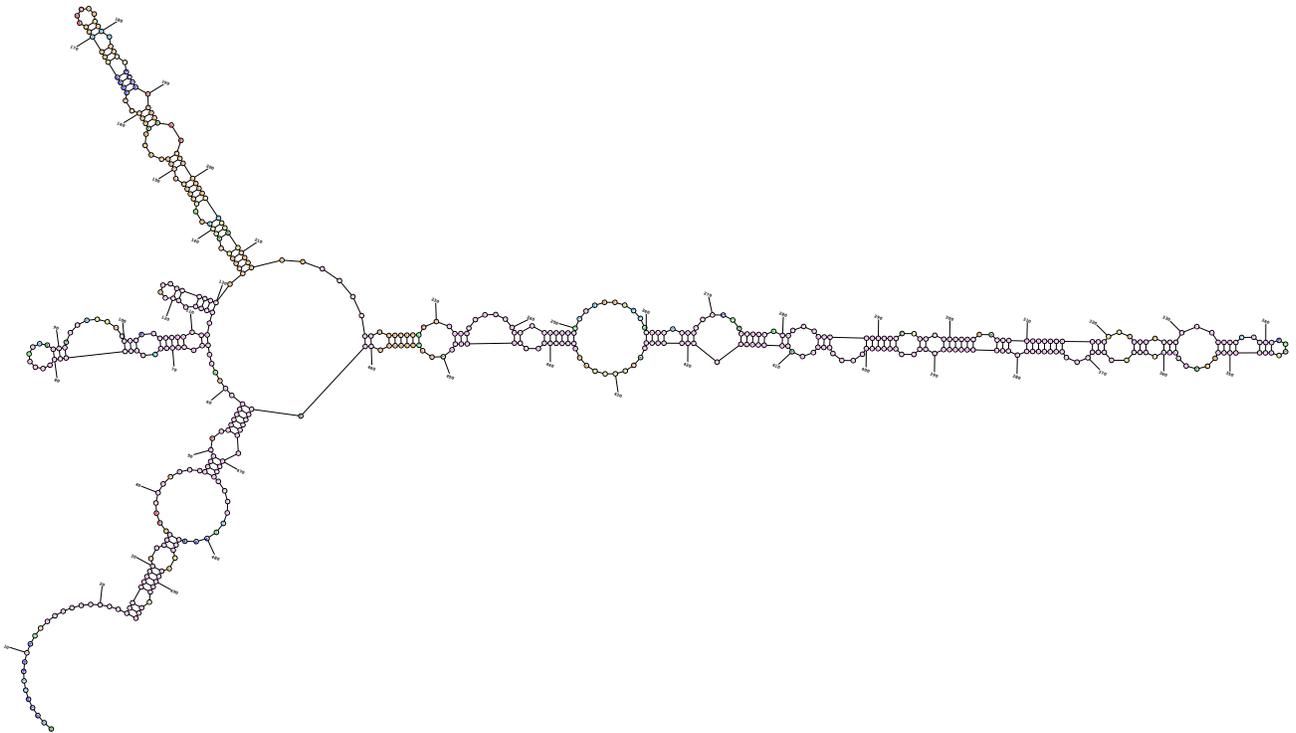
Supplemental Figure 3.5. Secondary structure predictions of LINC00152. A and B) The two lowest free-energy structures of LINC00152, predicted by RNAstructure.
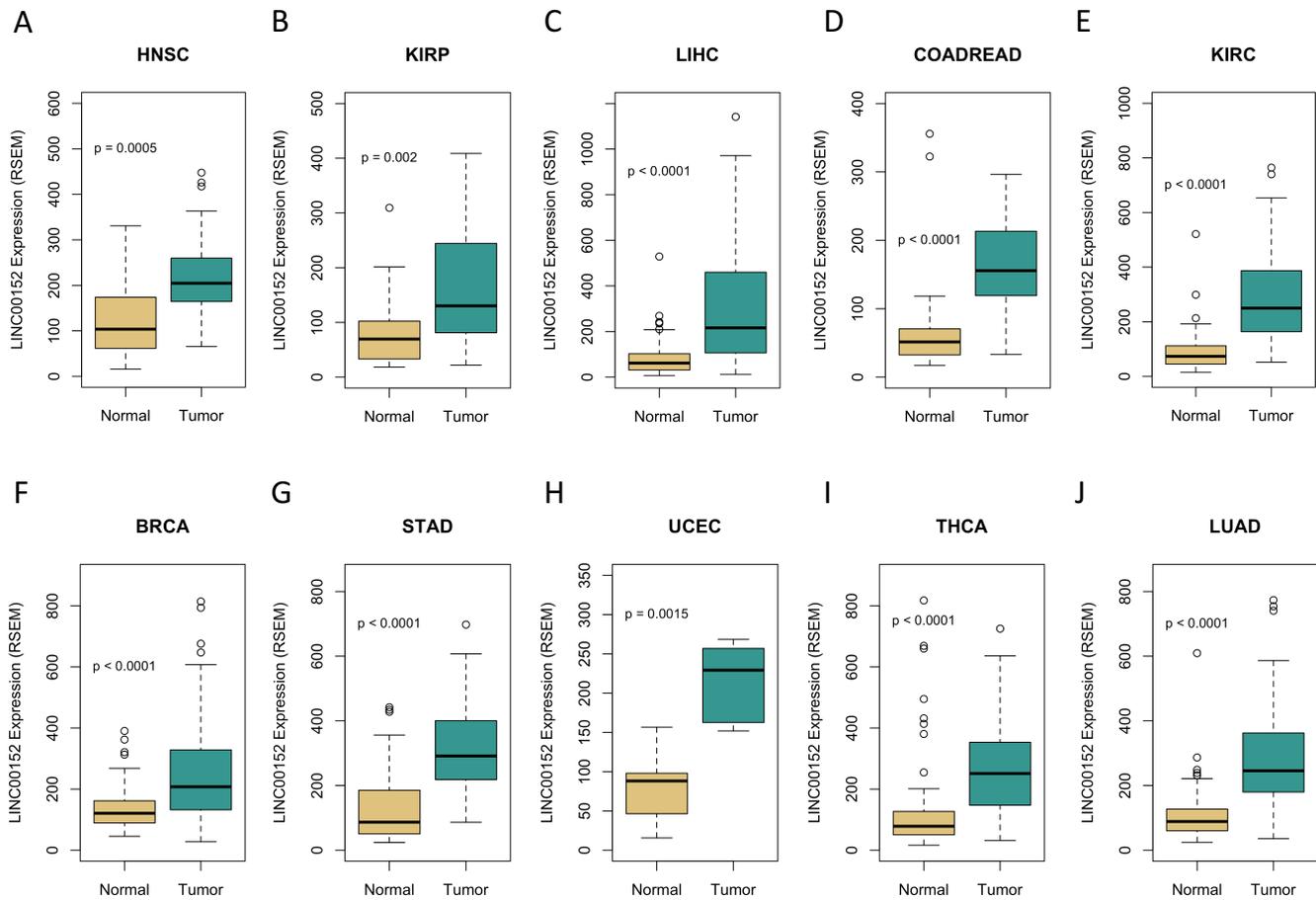
Figure 3.6. LINC00152 is upregulated in many cancer types. A – J) Expression (RSEM) of LINC00152 in tumors and matched normal tissue from the TCGA in head and neck squamous carcinoma, renal papillary tumor, hepatocellular carcinoma, colorectal carcinoma, renal clear cell carcinoma, breast invasive carcinoma, stomach adenocarcinoma, uterine carcinoma, thyroid carcinoma and lung adenocarcinoma, respectively.
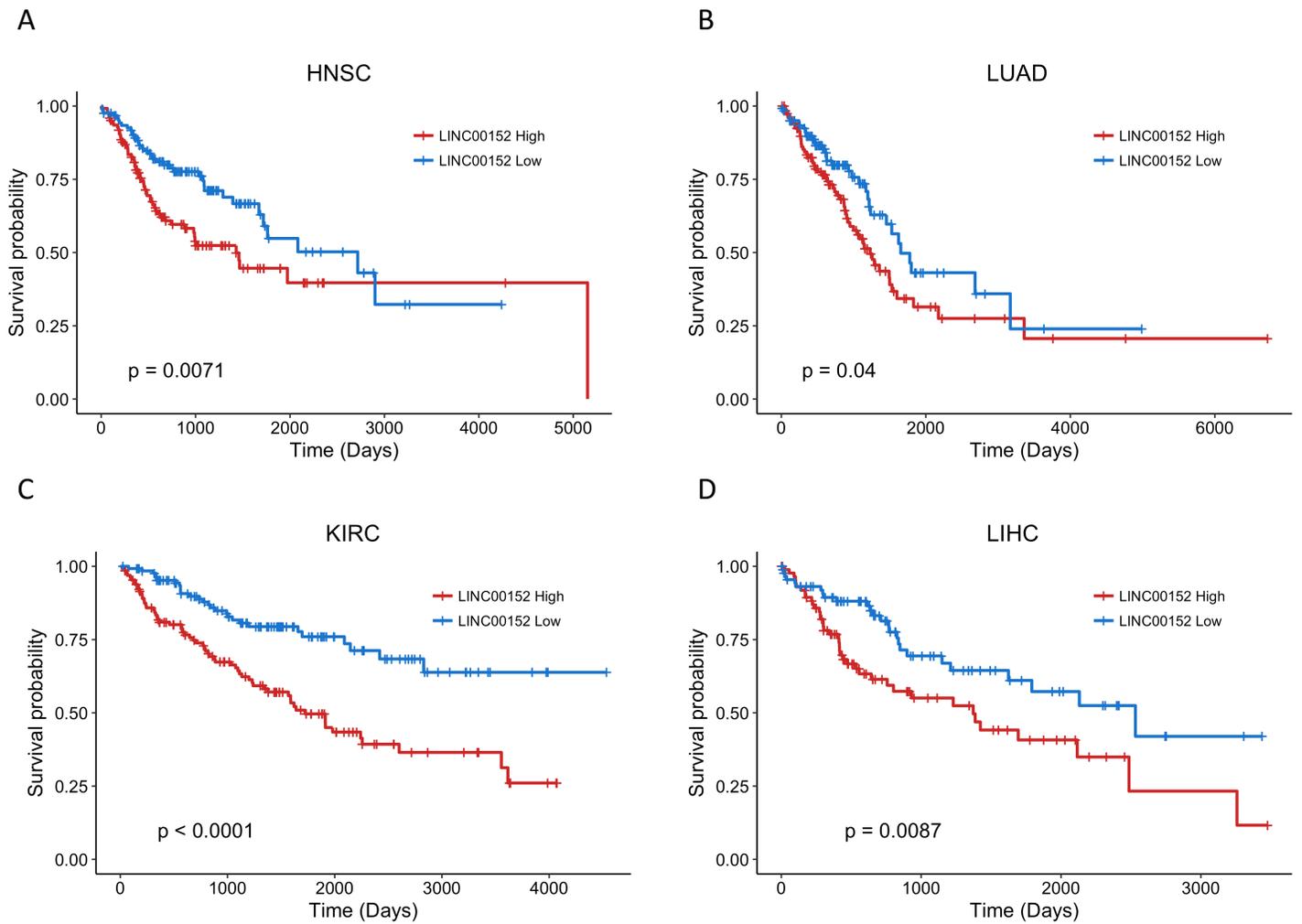
Figure 3.7 High levels of LINC00152 expression is associated with negative patient outcomes. A-D) Kaplan Meier plots of the highest LINC00152 expressing quartile and lowest LINC00152 expressing quartiles for head and neck squamous carcinoma, lung adenocarcinoma, renal clear cell carcinoma, and hepatocellular carcinoma, respectively.

A

KIRP

B

PAAD

C

LAML
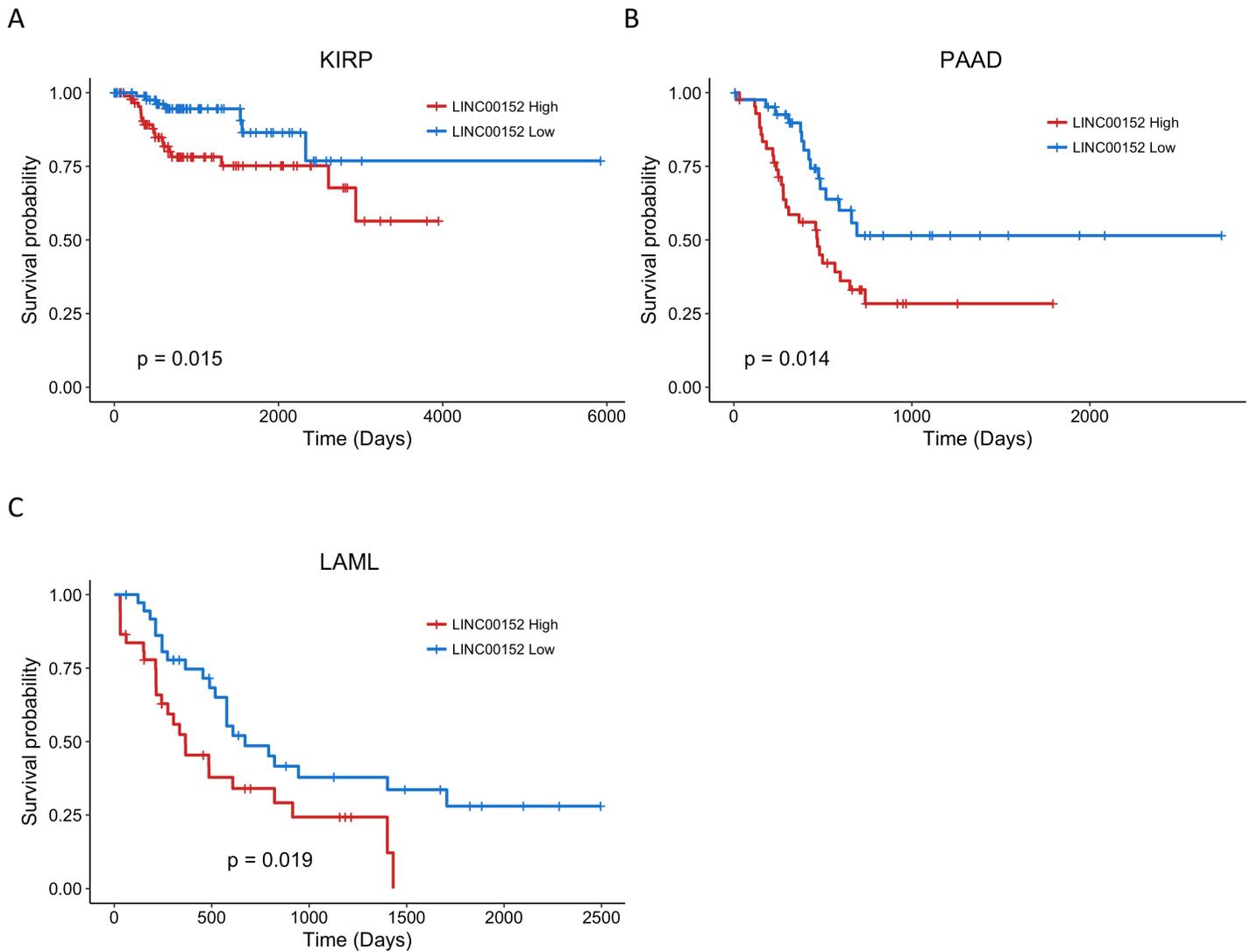
Supplemental Figure 3.6 High levels of LINC00152 expression is associated with negative patient outcomes. A-C) Kaplan Meier plots of the top third of LINC00152 expressing patients and the bottom third of LINC00152 expressing patients in renal papillary tumor, pancreatic adenocarcinoma, and acute myeloid leukemia, respectively.

Chapter 4

Discussion and Future Directions

Summary of Global Analysis of lncRNAs in Brain Cancer

We began this work to gain a comprehensive understanding of the extent to which lncRNAs are dysregulated in brain cancer as well as create a resource for future work investigating the role of lncRNAs in gliomas. To this end, we analyzed over 700 RNA-seq data from grade II, III and IV gliomas from the TCGA and normal brain tissue. Through our analysis, we identified over 2000 novel lncRNAs, many of which were differentially expressed in gliomas compared to normal brain tissue. Furthermore, we identified hundreds of lncRNAs that were preferentially upregulated or downregulated in specific LGG or GBM subtypes. In addition, we found similarities between the subtype specific lncRNAs in mesenchymal GBMs and IDH wt LGGs. Using multivariate survival analysis, we created a survival algorithm based on the expression of 64 lncRNAs that was capable of separating LGG patients into two distinct prognostic groups in both a test and validation patient data sets. We also identified lncRNAs that were associated with prognosis in specific GBM subtypes. Taken together, this work provides a unique and valuable resource that can aid in prioritizing which lncRNAs most likely play a role in gliomas and should be studied in more detail. However, there are still many improvements that can be made to help identify the lncRNAs that are most likely to be involved in glioma oncogenic phenotypes as well as create a more powerful survival algorithm. I will discuss several of these changes as well as the role of computational predictive algorithms in patient care in the following sections.

lncRNA Resource for Brain Tumors

One of the big contributions of our analysis of lncRNAs in GBMs and LGGs is that is creates a comprehensive list of lncRNAs that are associated with many GBM and LGG specific subtypes and genomic alterations. Previous studies have performed global meta-analyses that sought to identify lncRNAs deregulated in most of the cancer types with RNA-seq data from the TCGA [1,2]. Although these resources have provided a great resource for studying lncRNAs in cancer biology broadly, they are not as beneficial for studying lncRNAs in brain tumors for several reasons. Firstly, unlike many of the other tumors analyzed by the TCGA, GBMs and LGGs do not have many matched normal tissue RNA-seq data. This is due to the complication of biopsying normal brain tissue compared to other tissues such as liver and kidney. The lack of normal samples led to GBMs and LGGs being withheld from several of the analyses which relied on comparing to normal tissue. To address this issue, we included over 70 publicly available normal brain RNA-seq datasets to allow us to make these types of comparisons. Secondly, because the studies focused on many tumors, it prevented them from analyzing the association of lncRNAs with tumor specific characteristics such as subtypes and common somatic mutations. Our analysis fills this void by providing lists of lncRNAs associated with the aforementioned characteristics as well as list lncRNAs that were associated with patient survival. With these elements, the scientific community is now able to prioritize the lncRNAs that they study by examining their expressions in gliomas compared to normal tissue and whether or not they are associated with patient survival.

Survival Prediction in Cancers

In our study, we also designed a survival algorithm that can separate LGGs into two distinct survival classes based on the expression of a panel of 64 lncRNAs [3]. Although this application of lncRNAs expression and Cox regression in LGGs is novel, this is not the first

example of lncRNA or protein coding gene expression being used as a means to separate tumors into distinct subtypes or prognostic groups. Using transcriptomics to subdivide tumors is in many respects analogous to what the experimental pathology community has been doing for decades. Tumors are still categorized based on various attributes including cellular morphology, tumor cell heterogeneity, nuclear pleomorphism, mitotic activity and IHC (immunohistochemistry) staining, and clinical stage among other features. In fact, in many cases the histologic features are a direct reflection of the cell of origin or the genomic alterations that gave rise to the tumor (*i.e.* a reflection of the underlying transcriptomic alterations). Transcriptomic classification approaches have yet to be widely accepted, but transcriptome based analysis of breast cancer is gradually being accepted in regular clinical practice [4,5].

High throughput sequencing has revolutionized the way scientists conduct research and has broadened our understanding of fundamental biological processes. At its inception, genome sequencing technologies were extremely expensive, with the first human genome sequence costing nearly 3 billion dollars [6]. Since the onset of high throughput sequencing, the cost of genome sequencing has decreased exponentially and the throughput has increased sufficiently to make it available for commercial and clinical use. Indeed, in addition to commercial entities such as Foundation Medicine, many large comprehensive cancer centers now offer to sequence patient tumors to glean prognostic information as well as potential therapeutic targets. In response to the wealth of information that these efforts have created, basket trials have become a growing area of clinical trials, which are a type of clinical trial where treatment of a tumor is based on the presence of specific "actionable" mutations irrespective of the type of tumor they occur in [7,8]. In addition to prioritizing therapeutic options, the advances in sequencing have also aided in

expanding on the initial tumor classification efforts and provided more robust transcriptomic based biomarkers such as the PAM50 and OncoType Dx for breast cancer [9,10].

Predicting Survival in Gliomas

There are currently several genetic biomarkers that are used in the clinic to identify a glioma's susceptibility to conventional chemotherapy as well as glean prognostic information from the tumors genomic background. As mentioned earlier, IDH1 and IDH2 mutational status is one of the best predictors of overall survival in glioma patients, however the presence or absence of an IDH mutation does not change therapeutic treatment [11]. In addition, mutations in the TERT promoter has been shown to have clinical significance, as patients with mutational alterations in the TERT promoter have a worse prognosis in IDH 1/2 wt tumors. This can especially inform therapeutic strategies in low grade glioma patients that have favorable prognostic attributes. 1p19q co-deletion is pathognomonic for oligodendroglioma tumors and can be used to distinguish from other LGG histologic classes [12]. Moreover, clinical trials examining the role of procarbazine, CCNU/Lomustine and vincristine in addition to radiation therapy found that patients with 1p19q co-deletions had a better response rates compared to non-deleted tumors [13,14]. MGMT promoter methylation is also a well defined biomarker for gliomas that is used to predict a patient's response to standard of care chemotherapy, Temozolomide.

Beyond single gene or genomic site alterations, more recent work has found connections between clinical outcomes and transcriptional signatures. Indeed, the classical and mesenchymal GBM subtypes have increased response rates to intensive treatment, as opposed to proneural and neural subtypes [15]. However, this association has not been reproduced in an independent dataset, so it is unclear how strong the correlation between subtype and response to intensive

therapy actually is. Nonetheless, it illustrates the potential that transcriptional profiling of gliomas could play in future patient care. Furthermore, with the more widespread use of high throughput sequencing of patient tumors at academic cancer institutes, having better prediction methods to identify patients more at risk or more likely to respond to specific types of therapies will be extremely helpful. Our survival algorithm provides a preliminary example of the type of computational tools that could be used to identify more at risk glioma patients.

Improvements in the Survival Algorithm

Although our initial results show great promise, there are several changes and validation steps that should be done to increase its future clinical utility. One of the obvious modifications that could be made to our survival algorithm that would potentially improve its predictive value is to include mRNAs in addition to lncRNAs during the initial algorithm design. Many studies have used mRNA expression profiles to predict survival in a variety of tumors [16,17], including a study I participated in on LGG [18]. Including mRNAs in our survival prediction would most likely be beneficial as it would increase the initial pool of RNAs that are used to create the prognostic gene panel, allowing us to identify new prognostic genes. These new RNAs would hopefully capture some additional survival information that is not present in the current 64 lncRNA gene set, leading to a better, more robust survival model.

During the initial phase of designing our survival algorithm, we performed cox regression for all lncRNAs independently, and included the following variables into our multivariate analysis: tumor grade, gender, age, IDH mutational status and the expression for that particular lncRNA. We did not include the expression for all lncRNAs into one cox regression algorithm because increasing the amount of variables to over 2000 (total number of lncRNAs assayed), decreases Cox regression's ability to accurately identify individually prognostic lncRNAs [19].

As a consequence, although we identified 64 lncRNAs that consistently predicted patient survival, it is possible that many of these prognostic lncRNAs have expression patterns that strongly correlate or anti-correlate with each other, leading to unnecessary redundancy in the survival model. One way to simplify the algorithm would be to identify the groups of correlated and anti-correlated lncRNAs and take representative members from each correlation cluster to remove the redundancy. We could either select a single lncRNA from each correlation group that had the strongest association with survival (*i.e.* highest Cox coefficient) or we could create a composite score for each correlation group that takes into account the cox coefficients of all the lncRNAs within a given group.

Another modification that could be used to improve our survival algorithm would be to use machine learning to aid in the algorithm design. Machine learning is a type of data analysis that relies on a computer's ability to learn information about a given dataset without being explicitly told what to focus on. Machine learning has recently been gaining in popularity in the field of survival prediction, because of its ability to find hidden connections between desired outputs (i.e. survival) and a set of complex multi-variable inputs (*i.e.* gene expression, tumor grade, histological features, etc.). There are various machine learning methodologies, including Neural Networks (NN) and Support Vector Machines (SVM), which have been applied to predict patient survival in many cancer [20–23]. We could also employ SVM or NN in the generation of a new LGG survival model in conjunction with using an independent validation data set to test whether this methodology would create a more robust then our original 64 lncRNA survival algorithm.

Survival LncRNAs in Multiple Cancers

We designed our survival algorithm based on expression data from LGGs and have shown that it is capable of predicting survival in LGGs. However, we have not tested whether this survival algorithm is capable of predicting survival in other tumor types. Due to the heterogeneity within a given tumor type it is difficult to say whether our algorithm will be globally applicable to other cancers, but it will be interesting to see whether to what degree the LGG lncRNA panel is able to predict survival in tumors more broadly. Prior to testing our LGG lncRNA panel in other tumors it is extremely important to test the survival algorithm in a truly independent group of grade II and III glioma patients. Even though we initially validated the algorithm against a set of patients that were blinded from the initial algorithm design, the test patient set and the validation patient set were all from the TCGA cohort, which could bias the algorithm. After validation in an independent dataset, it will be interesting to see whether it can be used to predict survival in GBMs, as this tumor type is most similar to LGGs. Following this analysis, it will be interesting to see how well the LGG survival algorithm could be applied to other non-glial tumors. In addition to using the algorithm to separate patients into distinct survival cohorts, it would be interesting to compare lists of lncRNAs that are associated with patient prognosis across tumors to potentially identify key lncRNA oncogenes or tumor suppressors, similar to what we found with DRAIC and LINC00152 [24]. This type of analysis would be extremely beneficial for identifying which lncRNAs should be prioritized for experimental investigation and could be potential targets for future anti-tumor therapeutics.

Levels of Expression of LncRNA and Predictive Ability

For experimental studies it is best to focus on lncRNAs that have a reasonable level of expression, so that manipulation of lncRNA level (both down and up-regulation) can be carried out with ease. For this reason, I have plotted the cox coefficients of lncRNAs in predicting GBM

survival against their expression levels (Fig 4.1). Based on this chart and accompanying table (Table 4.1), the following lncRNAs should be prioritized for experimental manipulation in glioma cell lines in the future; RP11-521B24.5, CRNDE, RP11-571M6.7, CTD-3185P2.1, CTD-2336O2.1, AF127577.10, RP11-338K17.8, RP11-660L16.2, GS1-124K5.4 and LINC00663. To demonstrate the utility of this approach, we tested whether GS1-124K5.4, which is upregulated in both GBMs and LGGs compared to normal tissue and associated with survival in GBMs, expression affects any oncogenic phenotypes (Fig 4.2a and b). Indeed, knock down of GS1-124K5.4 led to a decrease in U87 cell growth and overexpression of exogenous GS1-124K5.4 resulted in increased cell proliferation (Fig 4.2c and d). These preliminary results highlight the benefits of using our lncRNA resource for prioritizing candidate lncRNAs for further study.

Summary of LINC00152 and GS1-1245K.4

From our original global analysis of lncRNAs in gliomas, I selected two lncRNAs, LINC00152 and GS1-124K5.4 to study in greater detail because both lncRNAs are significantly upregulated in GBMs and aggressive LGGs compared to normal brain tissue and high expression levels of each lncRNA is associated with negative patient outcomes. Furthermore, overexpression of LINC00152 or GS1-124K5.4 results in increasing invasion and cell growth, respectively, while knocking them down has the reciprocal effect. Initial secondary structure analysis of LINC00152 suggests that a stem-loop in its 3' end is partially involved in its invasive properties. Although our initial studies into the function of these two lncRNAs suggest that they might play key roles in gliomas, an understudied area of lncRNA research, more work is needed to fully understand how they function mechanistically in this aggressive tumor type.

Assessing the Oncogenic Role of LINC00152 and GS1-124K5.4 *in vivo*

Perhaps most important for cancer biology, stable up- or down-regulation of these two lncRNAs in U87 and A172 GBM cell lines should be tested in xenografts where the tumor cells are injected into the brains of nude mice to ascertain whether increase of these lncRNAs leads to a more malignant tumor. The cells will be marked by the stable expression of luciferase before injection. We can measure rate of growth of tumor volume, time to death (or sacrifice), angiogenesis, local invasion, distant metastases in these mice and determine whether increase of either of the two lncRNAs leads to a more malignant phenotype. If distant metastases are difficult to get from intracranial xenografts, we can inject the engineered cancer cells intravenously into the tail vein of nude mice and measure the extent of lung seeding of the tumors (by luciferin imaging in live mice, histology of lung sections and Q-PCR for human Alu DNA repeats after purification of mouse lung DNA).

FISH, RAP or RNA-IP and sequencing interacting RNAs

Although not a steadfast rule, a lot can be gleaned from a lncRNA's subcellular localization. As mentioned before, knowing where a lncRNA preferentially resides can help one prioritize potential functional pathways that a lncRNA effect. Using traditional cellular fractionation techniques, we have shown that LINC00152 is primarily located in the cytoplasm of U87 cells (Sup Fig 3.3). One issue with traditional cellular fractionation protocols is that fractionation is not always clear and nuclear proteins or RNAs can leak into the cytoplasmic fractions, leading to an incorrect interpretation of a lncRNAs localization. Furthermore, even if an RNA is correctly determined to be in the cytoplasm or nucleus, these types of fractionation techniques do not give detailed information on how they are localized within a given cellular compartment. Single-molecule RNA FISH (fluorescence in situ hybridization) is a technique that can be used to identify the location of individual RNA molecules [25]. Indeed, a study

performing single molecule FISH on lncRNAs identified 5 distinct localization patterns [26]. Two of the localization patterns were for lncRNAs primarily present in the cytoplasm, and the other was for lncRNAs present in both the cytoplasm and the nucleus. Interestingly, nuclear lncRNAs have 3 main localization patterns, either diffusely scattered throughout the nucleus, one or two large nuclear foci, or a few large foci in the nucleus with other small molecules dispersed throughout the nucleus [26]. Performing single-molecule RNA FISH on LINC00152 and GS1-124K5.4 will be very informative towards future comprehensive characterization efforts.

LncRNAs functions are mechanistically diverse, although most have been found to operate, in part, through base-pairing with DNA or RNA [27]. Identifying which DNA sequences or RNA transcripts a lncRNA interacts with is critically important for understanding how it functions. Although recently more appreciation has been directed toward lncRNA-RNA interactions, traditionally most of the studies characterizing lncRNA function was geared toward identifying where along the genome a lncRNA binds. In fact, many of the methodologies that are now used to find lncRNA RNA interaction partners are adaptations of techniques that were originally designed to identify interacting genomic DNA, such as ChIRP-seq (chromatin isolation by RNA precipitation) and RAP (RNA antisense purification) [28,29]. Now, several methods exist that identify RNA-RNA interactions globally or in a targeted manner towards a specific transcript's RNA interactome [30–32]. Indeed, this is the same methodology that was used to characterize the role of TINCR role in the cytoplasm and led to the discovery that it stabilizes several mRNAs through interacting with a 25-bp motif [32]. By performing targeted LINC00152 and GS1-124K5.4 pull-downs followed by RNA-seq to identify interacting RNAs, we will gain a better understanding of how both of these transcripts function.

Further Structure Function Elucidation of LINC00152 and GS1-124K5.4

In the early days on mammalian lncRNA research, there wasn't a huge focus on defining secondary structure elements of a particular lncRNA, due to the reliance on *in vitro* secondary structure mapping. Moreover, many recent studies blindly make deletion mutations in a lncRNA without giving weight to where the deletion boundaries might fall within the secondary structure, which can lead to faulty interpretations about the influence of a particular section of the lncRNAs on its function. More recently, newer technologies have been developed that map *in vivo* RNA secondary structure in a high throughput manner [33–37]. One drawback of these methodologies is that in order to get sequencing data for a specific RNA, it must be expressed at relatively high levels. Indeed, using PARIS data we were able to identify a secondary structure element in the highly expressed LINC00152, but could not for GS1-124K5.4, which is present at much lower levels compared to LINC00152 (Sup Fig 3.4) [31]. Furthermore, we could only identify secondary structure information for a single region of LINC00152 but not for the remaining 80% of its sequence. One way to get around this limitation would be to perform PARIS, or one of the other *in vivo* secondary structure mappings and then perform targeted sequence capturing using biotinylated oligos complementary to LINC00152 or GS1-124K5.4 pull-down and enrich secondary structure elements that are derived from those RNAs. This enrichment step should greatly increase the amount of sequencing reads that come from LINC00152 or GS1-124K5.4. By performing this type of secondary structure analysis, we should get a more precise and comprehensive map of LINC00152 or GS1-124K5.4 folding. This in turn improves our ability to make deletion mutant constructs and better probe structure function relationships in these lncRNAs.

Identifying LINC00152 and GS1-124K5.4 Interacting Proteins

One of the largest hurdles to characterizing a lncRNA's function is to identify interacting proteins. Nearly every example of mammalian lncRNAs have been demonstrated to function through interacting with some protein effector. Traditionally, interacting proteins have been identified through *in vitro* pull-down assays that rely on the use of a tagged or chemically labeled RNA [38–40]. Although these *in vitro* association assays have led to critical insights in lncRNA function, they can often lead to non-physiologic interactions or preferential pulling-down of highly abundant proteins [41,42]. More recently, several techniques have been developed to identify lncRNA interacting proteins *in vivo* that do not rely on the use of RNA tags or labeling [43,44]. Most of these methods use antisense nucleic acids to isolate RNA-protein complexes, either under native conditions or in denaturing conditions for samples that have been cross-linked [44]. Performing these types of unbiased *in vivo* interaction studies would help shed light on how LINC00152 and GS1-124K5.4 function. Furthermore, using more refined secondary structure information, the protein binding sites can be mapped using lncRNA deletion mutants and *in vitro* pull-down assays. Gaining a better appreciation of the LINC00152 and GS1-124K5.4 protein interaction repertoire will help improve our understanding out how these oncogenic lncRNAs are functioning in brain cancer.

Potential Therapeutic Applications of Targeting lncRNAs

In addition to using lncRNAs as biomarkers, recent work has speculated on the clinical utility of actually targeting lncRNAs as part of a comprehensive treatment regimen for patients [45–47]. There are several different technologies that could be used to target lncRNAs therapeutically, including siRNAs, molecules that bind lncRNAs or chemicals that can augment the expression of lncRNAs. siRNAs have already been used in clinical trials in various diseases to lower target gene expression [48,49]. LncRNAs are perfect candidates for siRNA based

therapies because most siRNAs target RNA levels, which is the final functional unit of a lncRNA, unlike proteins where a single mRNA can give rise to many proteins. This signal amplification may necessitate higher levels of RNA knockdown to achieve a clinical response, compared to lncRNAs. Modarresi *et al.* intracranially administered siRNA targeting the lncRNA BDNF-AS in mice and saw significant decrease in expression and increased neuronal outgrowth [50]. The main issue with siRNA based therapies is the difficulty of targeting a specific tissue, however recent advances in siRNA delivery systems have demonstrated that siRNAs can be delivered to specific tissues [51].

There are currently many compounds that are used in the clinical setting that target RNA-protein complexes. The most famous of which are antibiotics, such as streptomycin and tetracyclines, that target the bacterial ribosome [52,53]. Another example of small molecules used to target RNAs is a group of compounds that bind a crucial HIV RNA element, TAR (transactivation response element) [54]. These small molecules bind to TAR with high affinity and prevent it from interacting with its protein binding partner [54]. In addition to using compounds to directly bind a lncRNA, they could also be used to alter the expression of lncRNAs by targeting proteins that regulate their stability or expression. An example of such a compound is I-BET151, which is a selective BET (bromodomain and extraterminal) bromodomain inhibitor [55]. BETs such as BRD2, BRD3 and BRD4, which are upregulated in GBMs, play a critical role in regulating gene expression [56]. I-BET151 strongly inhibits GBM cell growth both *in vitro* and *in vivo,* by affecting the expression of many lncRNAs including HOTAIR [57]. Indeed, re-expressing HOTAIR is sufficient to restore cell growth during I-BET151 treatment [57]. These studies illustrate various methodologies that can be used to target

lncRNAs in cancer and future research into designing newer lncRNA targeted small molecules

or siRNAs could increase our therapeutic arsenal to target cancer and other diseases.

References

1.   Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet [Internet]. 2015 Mar;47(3):199–208. Available from: http://dx.doi.org/10.1038/ng.3192

2.   Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, et al. Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. Cancer Cell. 2015 Jan;28(4):529–40.

3.   Reon BJ, Anaya J, Zhang Y, Mandell J, Purow B, Abounader R, et al. Expression of lncRNAs in Low-Grade Gliomas and Glioblastoma Multiforme: An In Silico Analysis. PLOS Med [Internet]. 2016 Dec 6;13(12):e1002192. Available from: http://dx.doi.org/10.1371%2Fjournal.pmed.1002192

4.   McVeigh TP, Hughes LM, Miller N, Sheehan M, Keane M, Sweeney KJ, et al. The impact of Oncotype DX testing on breast cancer management and chemotherapy prescribing patterns in a tertiary referral centre. Eur J Cancer [Internet]. 2014 Nov 19;50(16):2763–70. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4204201/

5.   Carlson JJ, Roth JA. The impact of the Oncotype Dx breast cancer assay in clinical practice: a systematic review and meta-analysis. Breast Cancer Res Treat [Internet]. 2013 Aug 24;141(1):13–22. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3902996/

6.   Gyles C. The DNA revolution. Can Vet J [Internet]. 2008 Aug;49(8):745–6. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2465779/

7.   Hyman DM, Puzanov I, Subbiah V, Faris JE, Chau I, Blay J-Y, et al. Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations. N Engl J Med [Internet]. 2015 Aug 20;373(8):726–36. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4971773/

8.   Hyman DM, Solit DB, Arcila ME, Cheng D, Sabbatini P, Baselga J, et al. Precision medicine at Memorial Sloan Kettering Cancer Center: clinical next-generation sequencing enabling next-generation targeted therapy trials. Drug Discov Today [Internet]. 2015 Dec 28;20(12):1422–8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4940024/

9.   Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. N Engl J Med [Internet]. 2004 Dec 30;351(27):2817–26. Available from: http://dx.doi.org/10.1056/NEJMoa041588

10.   Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. J Clin Oncol [Internet]. 2009 Mar 10;27(8):1160–7. Available from: http://ascopubs.org/doi/abs/10.1200/JCO.2008.18.1370

11.   Liu A, Hou C, Chen H, Zong X, Zong P. Genetics and Epigenetics of Glioblastoma: Applications and Overall Incidence of IDH1 Mutation. Front Oncol [Internet]. 2016 Jan 29;6:16. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4731485/

12.   Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol [Internet]. 2016;131(6):803–20. Available from: http://dx.doi.org/10.1007/s00401-016-1545-1

13.   van den Bent MJ, Brandes AA, Taphoorn MJB, Kros JM, Kouwenhoven MCM, Delattre J-Y, et al. Adjuvant Procarbazine, Lomustine, and Vincristine Chemotherapy in Newly Diagnosed Anaplastic Oligodendroglioma: Long-Term Follow-Up of EORTC Brain Tumor Group Study 26951. J Clin Oncol [Internet]. 2013 Jan 20;31(3):344–50. Available from: http://ascopubs.org/doi/abs/10.1200/JCO.2012.43.2229

14.   Taal W, van der Rijt CCD, Dinjens WNM, Sillevis Smitt PAE, Wertenbroek AAACM, Bromberg JEC, et al. Treatment of large low-grade oligodendroglial tumors with upfront procarbazine, lomustine, and vincristine chemotherapy with long follow-up: a retrospective cohort study with growth kinetics. J Neurooncol [Internet]. 2015;121(2):365–72. Available from: http://dx.doi.org/10.1007/s11060-014-1641-9

15.   Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell [Internet]. 2010 Dec 20;17(1):98–110. Available from: http://dx.doi.org/10.1016/j.ccr.2009.12.020

16.   Mallmann MR, Staratschek-Jox A, Rudlowski C, Braun M, Gaarz A, Wolfgarten M, et al. Prediction and prognosis: impact of gene expression profiling in personalized treatment of breast cancer patients. EPMA J [Internet]. 2010 Sep 20;1(3):421–37. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3405335/

17.   Glinsky G V, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest [Internet]. 2004 Mar 15;113(6):913–23. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC362118/

18.   Anaya J, Reon B, Chen W-M, Bekiranov S, Dutta A. A pan-cancer analysis of prognostic genes. Metpally R, editor. PeerJ [Internet]. 2015 Sep 16;3:e1499. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4815555/

19.   Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. Stat Med [Internet]. 1984 Apr 1;3(2):143–52. Available from: http://dx.doi.org/10.1002/sim.4780030207

20.   Chang S-W, Abdul-Kareem S, Merican AF, Zain RB. Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods. BMC Bioinformatics [Internet]. 2013;14(1):170. Available from: http://dx.doi.org/10.1186/1471-2105-14-170

21.   Siow-Wee C, Kareem SA, Kallarakkal TG, Merican AF, Abraham MT, Zain RB. Feature Selection Methods for Optimizing Clinicopathologic Input Variables in Oral Cancer Prognosis. Asia Pacific J Cancer Prev. 2011;12.

22.   Lisboa PJ, Taktak AFG. The Use of artificial neural networks in decision support in cancer: a systematic review. Neural Netw [Internet]. 2006;19. Available from: http://dx.doi.org/10.1016/j.neunet.2005.10.007

23.   Álvarez Menéndez L, de Cos Juez FJ, Sánchez Lasheras F, Álvarez Riesgo JA. Artificial neural networks applied to cancer detection in a breast screening programme. Math Comput Model [Internet]. 2010 Oct;52(7–8):983–91. Available from: http://www.sciencedirect.com/science/article/pii/S0895717710001378

24.   Sakurai K, Reon BJ, Anaya J, Dutta a. The lncRNA DRAIC/PCAT29 Locus Constitutes a Tumor-Suppressive Nexus. Mol Cancer Res [Internet]. 2015;13(5):828–38. Available from: http://mcr.aacrjournals.org/cgi/doi/10.1158/1541-7786.MCR-15-0016-T

25.   Dunagin M, Cabili MN, Rinn J, Raj A. Visualization of lncRNA by Single-Molecule Fluorescence In Situ Hybridization BT - Nuclear Bodies and Noncoding RNAs: Methods and Protocols. In: Nakagawa S, Hirose T, editors. Methods Mol Biol [Internet]. New York, NY: Springer New York; 2015. p. 3–19. Available from: http://dx.doi.org/10.1007/978-1-4939-2253-6_1

26.   Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biol [Internet]. 2015;16(1):20. Available from: http://dx.doi.org/10.1186/s13059-015-0586-4

27.   Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet [Internet]. 2014 Jan;15(1):7–21. Available from: http://dx.doi.org/10.1038/nrg3606

28.   Chu C, Qu K, Zhong F, Artandi SE, Chang HY. Genomic maps of lincRNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell [Internet]. 2011 Nov 18;44(4):667–78. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3249421/

29.   Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X-chromosome. Science [Internet]. 2013 Aug 16;341(6147):1237973. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3778663/

30.   Sharma E, Sterne-Weiler T, O'Hanlon D, Blencowe BJ. Global Mapping of Human RNA-

RNA Interactions. Mol Cell [Internet]. 2016 May 19;62(4):618–26. Available from: http://www.sciencedirect.com/science/article/pii/S109727651630106X

31. Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, et al. RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell [Internet]. 2016 May 19;165(5):1267–79. Available from: http://www.sciencedirect.com/science/article/pii/S0092867416304226

32. Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K, et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature. 2013 Jan;493(7431):231–5.

33. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung J-W, et al. Structural imprints in vivo decode RNA regulatory mechanisms. Nature [Internet]. 2015 Mar 26;519(7544):486–90. Available from: http://dx.doi.org/10.1038/nature14263

34. Sugimoto Y, Vigilante A, Darbo E, Zirra A, Militti C, D'Ambrogio A, et al. hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. Nature [Internet]. 2015 Mar 26;519(7544):491–4. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4376666/

35. Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. RNA SHAPE analysis in living cells. Nat Chem Biol [Internet]. 2013 Jan 25;9(1):18–20. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3706714/

36. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature [Internet]. 2014 Jan 30;505(7485):701–5. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3966492/

37. Flynn RA, Zhang QC, Spitale RC, Lee B, Mumbach MR, Chang HY. Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. Nat Protoc [Internet]. 2016 Feb 14;11(2):273–90. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4896316/

38. McHugh CA, Russell P, Guttman M. Methods for comprehensive experimental identification of RNA-protein interactions. Genome Biol [Internet]. 2014 Jan 27;15(1):203. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4054858/

39. Srisawat C, Engelke DR. Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. RNA [Internet]. 2001 Apr;7(4):632–41. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1370116/

40. Lee HY, Haurwitz RE, Apffel A, Zhou K, Smart B, Wenger CD, et al. RNA–protein analysis using a conditional CRISPR nuclease. Proc Natl Acad Sci [Internet]. 2013 Apr 2;110(14):5416–21. Available from: http://www.pnas.org/content/110/14/5416.abstract

41. Butter F, Scheibe M, Mörl M, Mann M. Unbiased RNA–protein interaction screen by

quantitative proteomics. Proc Natl Acad Sci [Internet]. 2009 Jun 30;106(26):10626–31. Available from: http://www.pnas.org/content/106/26/10626.abstract

42.   Zielinski J, Kilk K, Peritz T, Kannanayakal T, Miyashiro KY, Eiríksdóttir E, et al. In vivo identification of ribonucleoprotein-RNA interactions. Proc Natl Acad Sci [Internet]. 2006 Jan 31;103(5):1557–62. Available from: http://www.pnas.org/content/103/5/1557.abstract

43.   Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, et al. Systematic Discovery of Xist RNA Binding Proteins. Cell [Internet]. 2015 Apr 9;161(2):404–16. Available from: http://www.sciencedirect.com/science/article/pii/S0092867415003128

44.   McHugh CA, Chen C-K, Chow A, Surka CF, Tran C, McDonel P, et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature [Internet]. 2015 May 14;521(7551):232–6. Available from: http://dx.doi.org/10.1038/nature14443

45.   Parasramka MA, Maji S, Matsuda A, Yan IK, Patel T. Long non-coding RNAs as novel targets for therapy in hepatocellular carcinoma. Pharmacol Ther [Internet]. 2016 May;161:67–78. Available from: http://www.sciencedirect.com/science/article/pii/S0163725816300122

46.   Li Z, Rana TM. Therapeutic targeting of microRNAs: current status and future challenges. Nat Rev Drug Discov [Internet]. 2014 Aug;13(8):622–38. Available from: http://dx.doi.org/10.1038/nrd4359

47.   Fatemi RP, Velmeshev D, Faghihi MA. De-repressing LncRNA-Targeted Genes to Upregulate Gene Expression: Focus on Small Molecule Therapeutics. Mol Ther Nucleic Acids [Internet]. 2014 Nov 18;3(11):e196. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4461991/

48.   Tabernero J, Shapiro GI, LoRusso PM, Cervantes A, Schwartz GK, Weiss GJ, et al. First-in-Humans Trial of an RNA Interference Therapeutic Targeting VEGF and KSP in Cancer Patients with Liver Involvement. Cancer Discov [Internet]. 2013 Apr 10;3(4):406 LP-417. Available from: http://cancerdiscovery.aacrjournals.org/content/3/4/406.abstract

49.   Schultheis B, Strumberg D, Santel A, Vank C, Gebhardt F, Keil O, et al. First-in-Human Phase I Study of the Liposomal RNA Interference Therapeutic Atu027 in Patients With Advanced Solid Tumors. J Clin Oncol [Internet]. 2014 Dec 20;32(36):4141–8. Available from: http://ascopubs.org/doi/abs/10.1200/JCO.2013.55.0376

50.   Modarresi F, Faghihi MA, Lopez-Toledano MA, Fatemi RP, Magistri M, Brothers SP, et al. Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. Nat Biotech [Internet]. 2012 May;30(5):453–9. Available from: http://dx.doi.org/10.1038/nbt.2158

51.   Wittrup A, Lieberman J. Knocking down disease: a progress report on siRNA therapeutics. Nat Rev Genet [Internet]. 2015 Sep;16(9):543–52. Available from: http://dx.doi.org/10.1038/nrg3978

52.   Connell SR, Tracz DM, Nierhaus KH, Taylor DE. Ribosomal Protection Proteins and
        Their Mechanism of Tetracycline Resistance. Antimicrob Agents Chemother [Internet].
        2003 Dec;47(12):3675–81. Available from:
        http://www.ncbi.nlm.nih.gov/pmc/articles/PMC296194/

53.   Poehlsgaard J, Douthwaite S. The bacterial ribosome as a target for antibiotics. Nat Rev
        Micro [Internet]. 2005 Nov;3(11):870–81. Available from:
        http://dx.doi.org/10.1038/nrmicro1265

54.   Stelzer AC, Frank AT, Kratz JD, Swanson MD, Gonzalez-Hernandez MJ, Lee J, et al.
        Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble.
        Nat Chem Biol [Internet]. 2011 Aug;7(8):553–9. Available from:
        http://dx.doi.org/10.1038/nchembio.596

55.   Nicodeme E, Jeffrey KL, Schaefer U, Beinke S, Dewell S, Chung C, et al. Suppression of
        inflammation by a synthetic histone mimic. Nature [Internet]. 2010 Dec
        23;468(7327):1119–23. Available from: http://dx.doi.org/10.1038/nature09589

56.   Pastori C, Daniel M, Penas C, Volmar C-H, Johnstone AL, Brothers SP, et al. BET
        bromodomain proteins are required for glioblastoma cell proliferation. Epigenetics
        [Internet]. 2014 Apr 1;9(4):611–20. Available from:
        http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4121371/

57.   Pastori C, Kapranov P, Penas C, Peschansky V, Volmar C-H, Sarkaria JN, et al. The
        Bromodomain protein BRD4 controls HOTAIR, a long noncoding RNA essential for
        glioblastoma proliferation. Proc Natl Acad Sci [Internet]. 2015 Jul 7;112(27):8326–31.
        Available from: http://www.pnas.org/content/112/27/8326.abstract

**GBM Cox Coefficient vs Expression Volcano Plot**
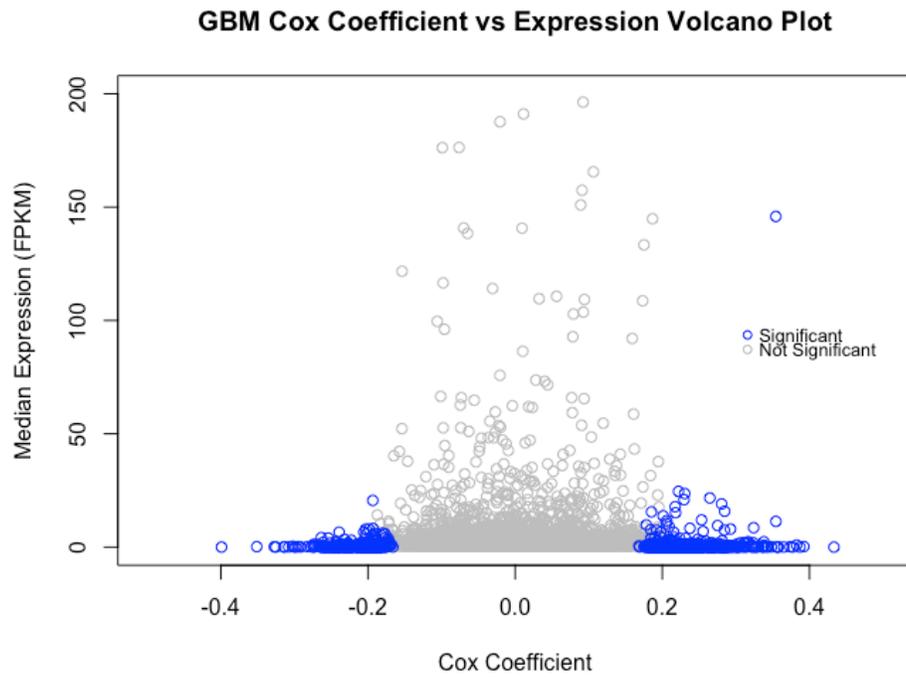
Figure 4.1. Volcano plot of lncRNAs separating them based on expression (FPKM) and cox coefficient. Blue points are those that are statistically significant (p value < 0.05).

lncRNA	Cox Coefficient	Median Expression (FPKM)
AGAP2-AS1	0.3544	11.3359
RP11-571M6.7	0.3543	145.893
RP11-74C13.4	0.3241	8.564445
CRNDE	0.2847	15.824
RP11-542C16.1	0.2927	7.973685
RP11-792A8.4	0.2806	19.0665
LINC01503	0.2752	9.52096
RP11-762H8.4	0.2844	7.49609
Cabili_2815	-0.2393	6.59275
CADM3-AS1	0.2646	21.677
RP11-334C17.6	-0.2635	4.20283
RP11-6N17.4	0.2533	12.00405
SOX21-AS1	-0.254	3.99134
LGG_LOC027071	0.2554	6.949825
AC068580.6	0.2316	3.85752
RP11-809N8.2	0.2375	8.292035
MIR497HG	0.2174	17.8172
RP11-359E10.1	0.2234	3.88899
SEMA6A-AS1	-0.2029	5.64052
CTB-31O20.9	-0.19873	7.971485
CTD-2336O2.1	0.222	24.58065
RP11-361L15.3	0.2295	20.9118
GCSHP3	-0.20565	6.592755
AC006116.24	-0.203	7.947185
LINC01057	0.2161	4.354855
RP11-1143G9.4	0.2304	23.64535
CTD-3185P2.1	0.206	11.67795
LGG_LOC040572	0.218	15.16545
GS1-124K5.4	0.2055	10.4736
RP11-27M24.1	0.2155	6.563745
RP11-644F5.11	0.198	6.48367
RP11-458J1.1	0.1897	4.3336
RP11-841O20.2	0.2092	9.061345
RP11-309L24.2	0.1844	3.796985
CTB-51J22.1	0.1967	4.81081
LINC00663	-0.1923	5.18596
STARD7-AS1	-0.1799	6.149375
RP11-660L16.2	-0.193	8.212445
RP4-665J23.1	0.1923	7.515495
RP11-338K17.8	0.2011	13.7827
GBM_LOC047769	-0.17648	4.19428
CTD-2583A14.8	-0.1725	3.719905
AF127577.10	0.1853	15.5451
CTD-2303H24.2	-0.1812	5.208445
BMS1P20	-0.1767	5.78431
ARSD-AS1	0.1782	9.81168
MIR22HG	0.1841	7.705755
RP5-1021I20.5	-0.1944	4.985235
RP11-521B24.5	-0.1935	20.5905
GBM_LOC002313	-0.19121	5.9063

Table 4.1. List of top lncRNAs associated with GBM patient prognosis that do not have low expression levels.
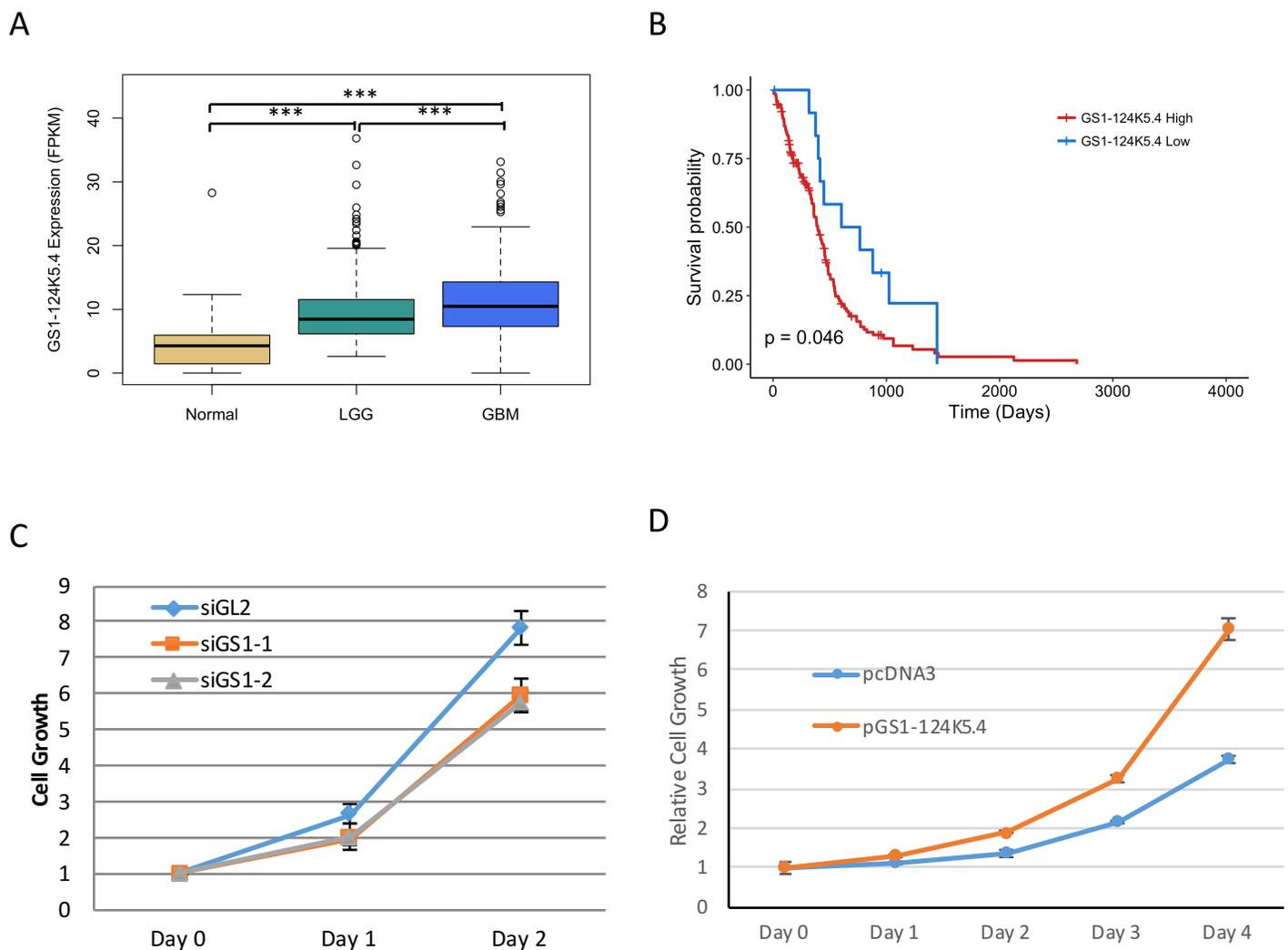
Figure 4.2. GS1-124K5.4 upregulated in GBMs and LGGs, associated with patient prognosis in GBMs and regulates U87 cell growth. A) Boxplot of LINC00152 expression in GBM, LGGs and normal brain tissue. B) Kaplan Meier plot of GBM patients comparing the top eight percentile of GS1-124K5.4 expressing patients with the remaining patients. C and D) U87 cell growth following knockdown or overexpression of GS1-124K5.4, respectively. *** p-value < 0.0001

Appendix

Scientific Contributions in other Studies

In addition to the work that I have outlined in the above chapters, I have also made contributions to following scientific work.

Mueller AC[1], Cichewicz MA[1], Dey BK[1], Layer R[1], **Reon BJ[1]**, Gagan JR[1], Dutta A[2]. MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis. Mol Cell Biol. 2015 Feb;35(3):498-513

- I performed microarray analysis of MUNC knockdown in growth media and differentiation media in C2C12 cells. I identified differentially expressed gene clusters associated cell growth and muscle differentiation.

Sakurai K[1], **Reon BJ[1]**, Anaya J[1], Dutta A[2]. The lncRNA DRAIC/PCAT29 Locus Constitutes a Tumor-Suppressive Nexus. Mol Cancer Res. 2015 May;13(5):828-38

- I performed the initial bioinformatics analysis that led to the identification of the tumor suppressive lncRNA, DRAIC.

Anaya J[1], **Reon B[2]**, Chen WM[3], Bekiranov S[2], Dutta A[2]. A pan-cancer analysis of prognostic genes. PeerJ. 2016 Feb 16;3:e1499.

- I helped analyze the RNA-seq data for LGGs which was used to find survival genes in LGGs.

Reon BJ[1], Dutta A[2]. Biological Processes Discovered by High-Throughput Sequencing. Am J Pathol. 2016 Apr;186(4):722-32.

- Co-wrote review article with Dr. Anindya Dutta

MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis.

Mueller AC[1], Cichewicz MA[1], Dey BK[1], Layer R[1], Reon BJ[1], Gagan JR[1], Dutta A[2].

Abstract

An in silico screen for myogenic long noncoding RNAs (lncRNAs) revealed nine lncRNAs that are upregulated more than 10-fold in myotubes versus levels in myoblasts. One of these lncRNAs, MyoD upstream noncoding (MUNC, also known as DRR(eRNA)), is encoded 5 kb upstream of the transcription start site of MyoD, a myogenic transcription factor gene. MUNC is specifically expressed in skeletal muscle and exists as in unspliced and spliced isoforms, and its 5' end overlaps with the cis-acting distal regulatory region (DRR) of MyoD. Small interfering RNA (siRNA) of MUNC reduced myoblast differentiation and specifically reduced the association of MyoD to the DRR enhancer and myogenin promoter but not to another MyoD-dependent enhancer. Stable overexpression of MUNC from a heterologous promoter increased endogenous MyoD, Myogenin, and Myh3 (myosin heavy chain, [MHC] gene) mRNAs but not the cognate proteins, suggesting that MUNC can act in trans to promote gene expression but that this activity does not require an induction of MyoD protein. MUNC also stimulates the transcription of other genes that are not recognized as MyoD-inducible genes. Knockdown of MUNC in vivo impaired murine muscle regeneration, implicating MUNC in primary satellite cell differentiation in the animal. We also discovered a human MUNC that is induced during differentiation of myoblasts and whose knockdown decreases differentiation, suggesting an evolutionarily conserved role of MUNC lncRNA in myogenesis. Although MUNC overlaps with the DRR enhancer, our results suggest that MUNC is not a classic cis-acting enhancer RNA (e-RNA) acting exclusively by stimulating the neighboring MyoD gene but more like a promyogenic lncRNA that acts directly or indirectly on multiple promoters to increase myogenic gene expression.

The lncRNA DRAIC/PCAT29 Locus Constitutes a Tumor-Suppressive Nexus.

Sakurai K[1], Reon BJ[1], Anaya J[1], Dutta A[2].

Abstract

Long noncoding RNAs (lncRNA) are emerging as major regulators of cellular phenotypes and implicated as oncogenes or tumor suppressors. Here, we report a novel tumor-suppressive locus on human chromosome 15q23 that contains two multiexonic lncRNA genes of 100 kb each: DRAIC (LOC145837) and the recently reported PCAT29. The DRAIC lncRNA was identified from RNA-seq data and is downregulated as prostate cancer cells progress from an androgen-dependent (AD) to a castration-resistant (CR) state. Prostate cancers persisting in patients after androgen deprivation therapy (ADT) select for decreased DRAIC expression, and higher levels of DRAIC in prostate cancer are associated with longer disease-free survival (DFS). Androgen induced androgen receptor (AR) binding to the DRAIC locus and repressed DRAIC expression. In contrast, FOXA1 and NKX3-1 are recruited to the DRAIC locus to induce DRAIC, and FOXA1 specifically counters the repression of DRAIC by AR. The decrease of FOXA1 and NKX3-1, and aberrant activation of AR, thus accounts for the decrease of DRAIC during prostate cancer progression to the CR state. Consistent with DRAIC being a good prognostic marker, DRAIC prevents the transformation of cuboidal epithelial cells to fibroblast-like morphology and prevents cellular migration and invasion. A second tumor-suppressive lncRNA PCAT29, located 20 kb downstream of DRAIC, is regulated identically by AR and FOXA1 and also suppresses cellular migration and metastasis. Finally, based on TCGA analysis, DRAIC expression predicts good prognosis in a wide range of malignancies, including bladder cancer, low-grade gliomas, lung adenocarcinoma, stomach adenocarcinoma, renal clear cell carcinoma, hepatocellular carcinoma, skin melanoma, and stomach adenocarcinoma.

A pan-cancer analysis of prognostic genes.

Anaya J[1], Reon B[2], Chen WM[3], Bekiranov S[2], Dutta A[2].

Abstract

Numerous studies have identified prognostic genes in individual cancers, but a thorough pan-cancer analysis has not been performed. In addition, previous studies have mostly used microarray data instead of RNA-SEQ, and have not published comprehensive lists of associations with survival. Using recently available RNA-SEQ and clinical data from The Cancer Genome Atlas for 6,495 patients, we have investigated every annotated and expressed gene's association with survival across 16 cancer types. The most statistically significant harmful and protective genes were not shared across cancers, but were enriched in distinct gene sets which were shared across certain groups of cancers. These groups of cancers were independently recapitulated by both unsupervised clustering of Cox coefficients (a measure of association with survival) for individual genes, and for gene programs. This analysis has revealed unappreciated commonalities among cancers which may provide insights into cancer pathogenesis and rationales for co-opting treatments between cancers.

Biological Processes Discovered by High-Throughput Sequencing.

Reon BJ[1], Dutta A[2].

Abstract

Advances in DNA and RNA sequencing technologies have completely transformed the field of genomics. High-throughput sequencing (HTS) is now a widely used and accessible technology that allows scientists to sequence an entire transcriptome or genome in a timely and cost-effective manner. Application of HTS techniques has led to many key discoveries, including the identification of long noncoding RNAs, microDNAs, a family of small extrachromosomal circular DNA species, and tRNA-derived fragments, which are a group of small non-miRNAs that are derived from tRNAs. Furthermore, public sequencing repositories provide unique opportunities for laboratories to parse large sequencing databases to identify proteins and noncoding RNAs at a scale that was not possible a decade ago. Herein, we review how HTS has led to the discovery of novel nucleic acid species and uncovered new biological processes during the course.