A first comprehensive analysis of Transcribed Ultra Conserved Regions uncovers important regulatory functions of novel non-coding transcripts in gliomas.

Myron K Gibert Jr[1], Ying Zhang[1], Shekhar Saha[1], Pawel Marcinkiewicz[1], Collin Dube[1], Kadie Hudson[1], Yunan Sun[1], Sylwia Bednarek[1], Bilhan Chagari[1], Aditya Sarkar[1], Christian Roig-Laboy[1], Natalie Neace[1], Karim Saoud[1], Initha Setiady[1], Farina Hanif[1], Pankaj Kumar[4], Benjamin Kefas, and Roger Abounader[1,2,3]

University of Virginia Departments of Microbiology, Immunology & Cancer Biology[1] Neurology[2], Cancer Center[3], and Department of Public Health Sciences and Bioinformatics Core[4], Charlottesville, VA, 22908, USA

† These authors contributed equally to this work.

Corresponding author: Roger Abounader, University of Virginia, PO Box 800168, Charlottesville VA 22908, USA, Phone: (434) 982-6634, Fax: 434-243-6843 E-mail: ra6u@virginia.edu.

ABSTRACT

Transcribed Ultra-Conserved Regions (TUCRs) represent a severely understudied class of putative non-coding RNAs (ncRNAs) that are 100% conserved across multiple species. We performed the first-ever analysis of TUCRs in glioblastoma (GBM) and low-grade gliomas (LGG). We leveraged large human datasets to identify the genomic locations, chromatin accessibility, transcription, differential expression, correlation with survival, and predicted functions of all 481 TUCRs, and identified TUCRs that are relevant to glioma biology. Of these, we investigated the expression, function, and mechanism of action of the most highly upregulated intergenic TUCR, uc.110, identifying it as a new oncogene. Uc.110 was highly overexpressed in GBM and LGG, where it promoted malignancy and tumor growth. Uc.110 activated the WNT pathway by upregulating the expression of membrane frizzled-related protein, by sponging the tumor suppressor miR-544. This pioneering study shows important roles for TUCRs in gliomas and provides an extensive database and novel methods for future TUCR research.

INTRODUCTION

Transcribed ultra-conserved regions (TUCRs) represent 481 unique transcribed RNA molecules that are "ultraconserved" across multiple species, including in the human, mouse (100%), rat (100%), dog (98%), and chicken (95%) genomes. [1] TUCR expression has been found to be highly deregulated in some cancers. Because of their ultra-conservation and their deregulation, it is believed that TUCRs may have important regulatory roles in cancer. [2-11] About ~90% of the genome is transcribed, but only ~2 percent of the transcriptome is then translated. The remainder of the transcriptome is made up of non-coding elements that serve key regulatory roles. Of these elements, long non-coding RNAs (lncRNAs) serve as important regulators of malignancy and potential therapeutic targets in cancer. [2, 12-19] Due to their size and lack of known associated protein products, it has been suggested that many TUCRs may function as lncRNAs.[2] The putative existence of "ultraconserved" lncRNAs is significant, as lncRNAs are typically poorly conserved as a class of molecules.[2] Very little is known about TUCRs. [2] In particular, the literature elucidating the expressions, functions, and mechanisms of action of TUCRs in glioblastoma (GBM) and low-grade glioma (LGG) is nonexistent. GBM and LGG represent over 80% of primary malignant brain tumors in humans, of which GBM is the deadliest, with a median survival of approximately 15 months.[20-28] Studying TUCRs in gliomas is therefore an untouched avenue for understanding novel oncogenic mechanisms and discovering new biomarkers and therapeutic targets.

In this study, we leveraged large human datasets to identify the genomic locations, chromatin accessibility, transcription, differential expression, correlation with survival, and predicted functions of all 481 TUCRs, and identified TUCRs that are relevant to glioma biology. Of these, we investigated the expression, function, and mechanism of action of the most highly upregulated intergenic TUCR, uc.110, identifying it as a new oncogene. Uc.110 was highly overexpressed in GBM and LGG, where it promoted malignancy parameters and tumor growth. Uc.110 activated the WNT pathway by upregulating the expression of membrane frizzled-related protein, by sponging the tumor suppressor miR-544. This pioneering work shows important roles for TUCRs in gliomas and provides an extensive database and novel methods for future TUCR research in any disease context.

RESULTS

TUCRs are encoded throughout the genome, resistant to variation, and actively transcribed.

We analyzed TUCR genomic locations published in Bejerano et al, [1] using hg38 genome coordinates lifted over from the provided hg19 coordinates. We found that some TUCRs are exonic and are contained within an exon of the "host" gene. Others are contained within an intron instead. Some TUCRs straddle a region that spans exonic and intronic regions of the host gene (exonic/intronic), and others are not contained within a known genetic element at all (intergenic) (Figure 1B). We manually annotated each TUCR using a combination of UCSC Genome Browser tracks, [29, 30] Quinlan Laboratory's bedtools, [31, 32] and TUCR genomic locations lifted over to hg38 from hg19. [1] We identified 45 exonic, 231 intronic, 68 intronic/exonic, and 137 intergenic TUCRs (Figure 1C). We found that TUCRs are located on all 22 numbered chromosomes and the X chromosome. There were no annotated TUCRs on the Y chromosome (chrY) or in the mitochondrial DNA (chrM) (Figure 1D). Detailed TUCR annotation information for each TUCR is provided in supplementary materials (Supplementary Master Table).

Since TUCRs are expected to be resistant to variation[2], we characterized the overlap of current dbSNP (build 156) single nucleotide polymorphism (SNP) annotations to the lifted over hg38 TUCR genomic coordinates. We found that TUCRs overlap with fewer SNPs than protein coding genes and non-coding RNAs, demonstrating that they are more resistant to variation (Figure 1E).

We also investigated the extent of TUCR transcription levels in comparison to transcription of known protein-coding and non-coding genes. To accomplish this, we first analyzed their spatial associations with markers for active chromatin (H3K4me3), active enhancers (H3K27ac), lncRNA transcription (RNA Pol.II) and open chromatin (ATAC-Seq). We determined the significance of the spatial relationships between these marks and TUCRs utilizing publicly available U87 CHIP- and ATAC-Seq datasets. Then, we compared these to TUCR intervals that were randomly shuffled to create a negative control, other classes of non-coding RNAs, and TUCRs subset by genomic annotation (Figure 1F and Supplementary Figure 1A). We found that TUCRs displayed a significant enrichment for all transcriptional activity markers over expected and compared to

control. The above data show that TUCRs are distributed throughout the genome, resistant to variation, and actively transcribed in GBM and LGG.

Figure 1

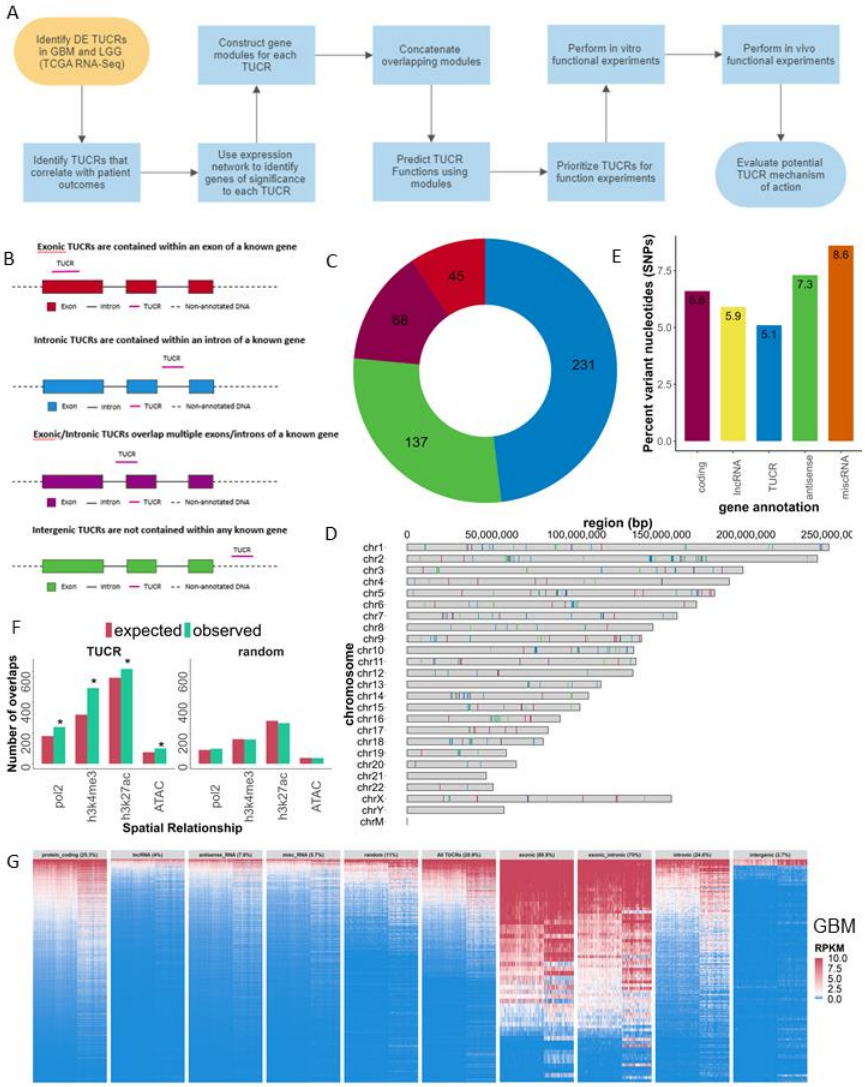


*Figure 1. Figure 1. Annotation, localization and expression of TUCRs in GBM and LGG. A) Experimental workflow for identifying and studying TUCRs of interest. B) TCGA analysis shows that TUCRs can be exonic (red), intronic (blue), exonic/intronic (purple) or intergenic (green). C) Circle graph showing the distribution of genomic annotations across all 481 TUCRs, with colors matching 1B. D) Karyoplot showing that TUCRs exist on all chromosomes except for the Y chromosome and mitochondrial DNA, vertical lines show TUCRs with colors matching 1B. E) Bar chart demonstrating that TUCRs are more resistant to single nucleotide variants (SNVs/SNPs) than other gene annotation categories. F) Bar chart showing that TUCRs are enriched for markers for open and active chromatin in U87 cells, suggesting that they represent transcriptionally active sites. Red bars represent chi-square expected overlaps, and teal bars represent observed values. G) Heatmap representing TUCR absolute expression (RPKM) across multiple gene annotations. Blue represents poorly expressed genes (<1 RPKM), White/Pink genes are moderately expressed (>=1 RPKM) and Red represents highly expressed genes (RPKM >=10). TUCRs demonstrate an expression profile that is most comparable with protein coding genes. * = p < 0.05*

TUCRs are highly expressed in GBM and LGG tumors.

TUCR expression has not been characterized in GBM or LGG before. We performed the first comprehensive analysis of TUCR expression in these cancers by comparing GBM (n = 166) and LGG (n = 505) tumor samples from the Cancer Genome Atlas (TCGA) [33] to their normal brain cortex counterparts in TCGA (n = 5) and the Genotype-Tissue Expression Database (GTex, n = 255). [34] We first analyzed absolute TUCR expression, as measured by reads-per-kilobase million (RPKM). The absolute expression, in GBM, of all TUCRs was compared to the expression of lncRNAs, coding genes, antisense RNAs, and small noncoding RNAs (< 200 nt length), and the expression of TUCRs separated by genomic annotation into exonic, intronic, exonic/intronic, and intergenic. All gene annotations were obtained using the CHESS gene catalog, which contains most Refseq and Ensembl genes, while also including a series of understudied novel genes.[35] Highly expressed genes are visualized via heatmap (>=10 RPKM, red) along with moderately (>=1 RPKM, white) and lowly expressed genes (<1 RPKM, blue). These analyses were repeated in LGG (Supplementary Figure 1B). The data show that intragenic TUCRs are expressed at magnitudes that are similar to those of protein coding genes in both GBM and LGG, while intergenic TUCRs demonstrate expression levels that are closer to those of lncRNAs. (Figure 1G).

TUCRs are deregulated in gliomas, and deregulation is associated with clinical outcomes.

We analyzed TCGA data and found that in addition to being highly expressed in gliomas, TUCRs are also highly deregulated. Of the 481 annotated TUCRs, we identified 87 that were upregulated and 67 that were downregulated in GBM (Figure 2A). We also identified 59 TUCRs that were upregulated and 53 TUCRs that were downregulated in LGG. (Figure 2B). Of the 154 deregulated TUCRs in GBM, 86 were also deregulated in LGG, a 56% overlap (Figure 2C). We then sought to determine whether deregulation of TUCR expression correlates with patient outcomes in LGG and GBM. For each of the 481 TUCRs, we generated a Kaplan-Meier plot tracking differences in survival for high expressing (upper quartile) and low expressing (lower quartile) groups. Of the TUCRs that are expressed in GBM TCGA RNA-Seq data, only 4 were correlated with survival in a statistically significant manner in both of our workflows ($p <= 0.05$, Supplementary Figure 3B). We considered that this low prevalence of survival-associated TUCRs in GBM was due to the poor overall survival of GBM patients (15 months). We also studied survival differences in LGG

patients, as they have a longer median survival (84 months). Of the TUCRs that are expressed in LGG TCGA RNA-Seq data, 93 were correlated with survival in both of our workflows (Figure 2D). Of these 93 TUCRs, 89 had a statistically significant Cox proportional hazard that was associated with a good prognosis, while 4 had one that was associated with a poor prognosis (Figure 2D). We have highlighted two TUCRs that represent a statistically significant correlation with good (uc.338, Figure 2E) or poor (uc.75, Figure 2F) prognosis using both methods. When separated by annotation category, intragenic TUCR deregulation have a greater association with patient outcomes than intergenic TUCRs (Supplementary Figure 3C).

TUCRs are coregulated with genes that have known functions.

We predicted TUCR function by identifying coregulated genes with known functions via weighted gene co-expression network analysis (WGCNA).[36] We aggregated the 42,644 genes in our dataset into 60 colored modules based on clustered gene ontology (GO) terms. Each of these modules contain genes with known functions, such as RNA binding, cell signaling, immune response, metabolic response, etc. These modules can also be used to identify genes that associate with clinical traits, such as the tumor tissue-type (Figure 2G). The data can also be used to predict gene function for novel genes. To do this, we also aggregated all 481 TUCRs into our modules. We were able to identify TUCRs that correlate with each of our 60 modules, with some having positive correlations and others negative. For example, TUCRs that exhibit a positive correlation with the “cyan” module (Supplementary Figures 3-2) could have a promoting effect in nucleic acid binding and regulation, while those that are negatively correlated with the “dark orange” module (Supplementary Figures 3-4) may have a negative effect on muscle metabolism. Since many different TUCRs show associations with different modules, and every module has at least one TUCR that associates with it, these results suggest that TUCRs may have a broad range of potential roles (Figure 2H).

Figure 2

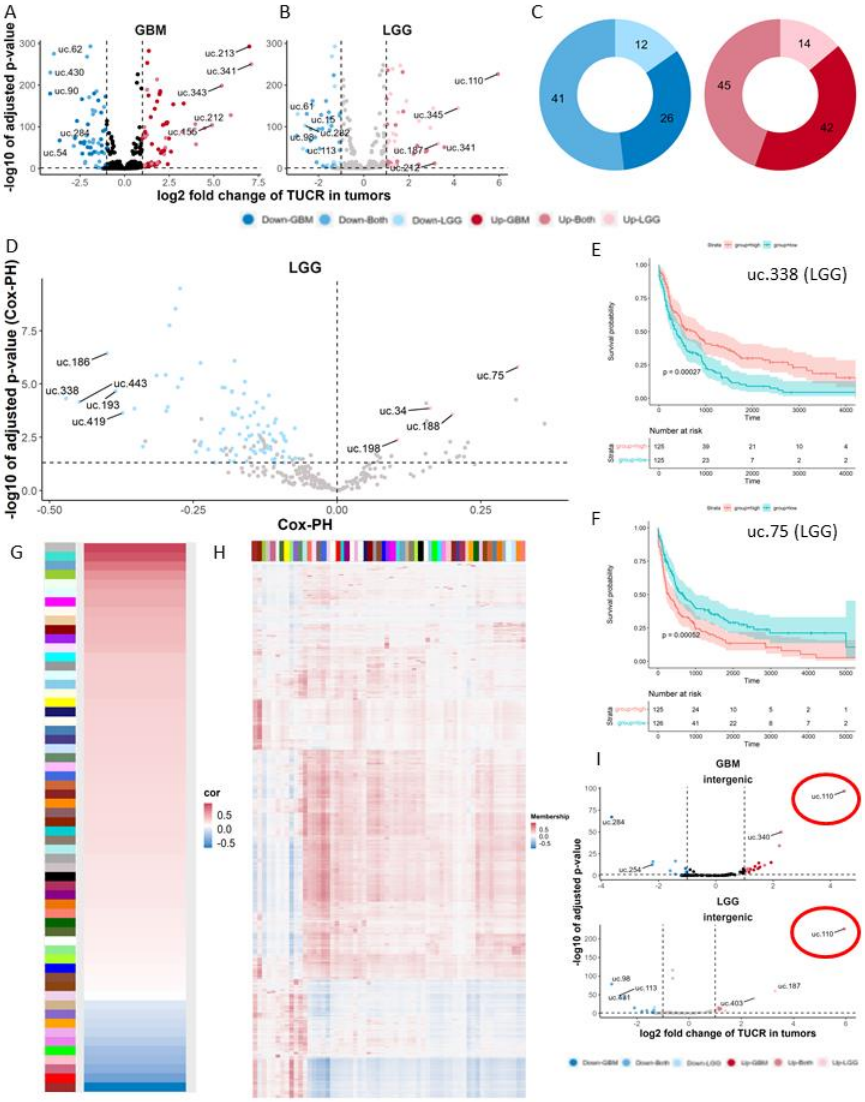


*Figure 2. Figure 2. TUCRs are deregulated and associated with patient outcomes in GBM and LGG and may have broad functional roles. All experiments were performed using TCGA GBM and LGG RNA-Seq data. A) Volcano plot showing that 87 TUCRs are upregulated >=2-fold (1-log2FC) and 67 are downregulated in GBM. Red dots are upregulated. Blue are downregulated. B) Volcano plot showing that 59 TUCRs are upregulated >=2-fold in LGG, and 53 are downregulated in LGG. C) Circle graph demonstrating that of the 154 deregulated TUCRs in GBM, 86 were also deregulated in LGG, a 56% overlap. Dark Red/Blue are TUCRs deregulated in GBM. Light Pink/Blue are TUCRs deregulated in LGG. Intermediate Red/Blue represent TUCRs deregulated in both. D) Volcano plot showing that several TUCRs are significantly associated with patient outcomes in LGG. Pink dots represent TUCRs significantly associated with poor prognosis. Blue dots represent TUCRs significantly associated with good prognosis. E) Kaplan-Meier showing that TUCR uc.338 is significantly associated with good prognosis. Red line represents the uc.338 high expression group. Teal line represents the uc.338 low expression group. F) Kaplan-Meier showing that uc.75 is significantly associated with poor prognosis. G) Heatmap showing module association with the tumor tissue trait. 42,644 genes, including the 481 TUCRs, were aggregated into color modules based on gene ontology (GO) terms. These modules demonstrate association with oncogenic (cor > 0, red) or tumor suppressor (cor < 0, blue) phenotypes. H) Heatmap showing that TUCRs demonstrate association with all 60 modules, suggesting broad potential functions. Red and Blue represent positive and negative correlations, respectively. I) Volcano plot showing that the uc.110 TUCR is the most upregulated TUCR in GBM and LGG and therefore may function as a novel oncogenic lncRNA. * = p < 0.05*

TUCR, uc.110, is highly upregulated in gliomas and is predicted to bind nucleic acids.

The expression and deregulation of intergenic TUCRs is of particular interest as they may represent novel lncRNAs due to their similar expression levels, genomic location, and lack of coding potential.[2] These TUCRs are also easier to study experimentally; they are often thousands of kilobases (kb) from the nearest protein-coding gene and likely function in a manner that is independent of a "host gene". Because of this, we focused on intergenic TUCRs for our experimental studies. Of the deregulated intergenic TUCRs in GBM in and LGG (Figure 2I), we found that uc.110 is the most upregulated; 22.3-fold in GBM and 61.4-fold in LGG (Figure 3A). It has near binary expression; it is rarely expressed at all in normal brain but is very highly expressed in GBM and LGG (Figure 3B). Due to its high expression, we hypothesized that this TUCR is functioning as an oncogene.

Since many TUCRs exist as a part of a larger transcript [2], we first determined the sequence of the uc.110 full transcript. We utilized machine learning and *de novo* transcript reassembly using TCGA and GTex RNA-seq data to reconstruct RNA-Seq transcripts in the absence of a reference genome (Supplementary Figure 4A). [35] We identified a 2,158 nt long RNA molecule that contains the 243 nucleotide (nt) uc.110 ultraconserved sequence (Figure 3C) as a novel transcript. We confirmed the existence of this transcript experimentally using PCR amplification and sequencing (Supplementary Figure 4B).

After identifying the sequence for the full uc.110 transcript (Supplementary Figure 4C), we utilized our WGCNA workflow to identify genes and modules (Figure 3D) that are significant to this transcript. Of note, each of the top 4 modules for uc.110 by module association represents genes that are involved in nucleic acid and protein binding (Supplementary Figures 3). These modules are all associated with an oncogenic phenotype (Figure 2G) and genes within these modules are positively coregulated with uc.110 (Figure 3E). Based on these findings, we hypothesized that uc.110 may be operating as an oncogenic RNA-binding molecule. We also performed similar analyses for all 481 TUCRs to identify potential functional roles for each TUCR in gliomas (Supplementary Figures 5).

Figure 3

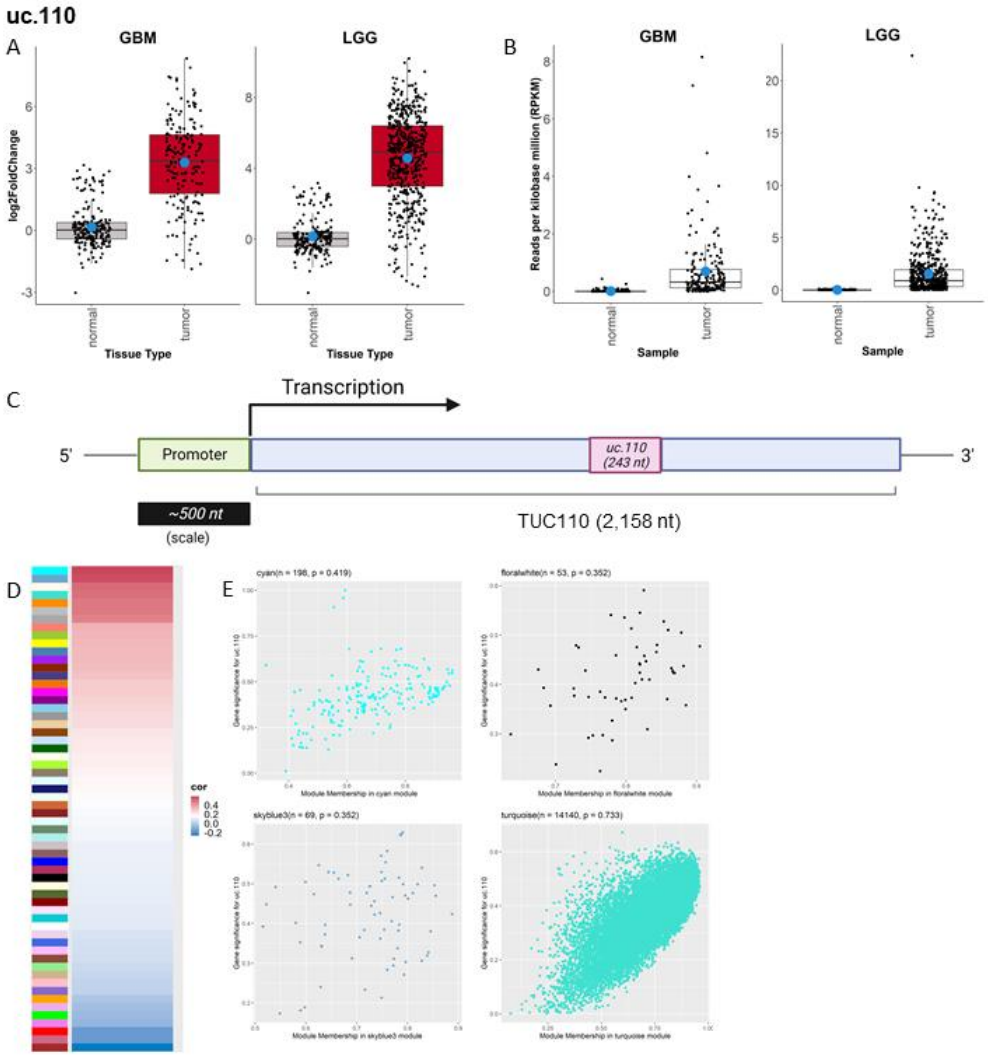


Figure 3. The uc.110 TUCR is the most upregulated intergenic TUCR in gliomas. A) Box- and dotplot showing that uc.110 is 30-fold upregulated in GBM and 60-fold upregulated in LGG based on TCGA data analysis. Red boxes represent upregulated TUCRs. Green boxes represent downregulated TUCRs. Gray boxes represent TUCRs that are not deregulated. B) Box- and dotplot showing that uc.110 is expressed in tumors but is poorly expressed in normal brain cortex based on TCGA and Gtex data analyses. C) Cartoon showing that uc.110 is a 243 nt region in a 2,158 nt transcript. D) Heatmap depicting uc.110 gene module association. Positive correlations are red, while negative correlations are blue, with weak correlations in white. uc.110 shows significant module membership for the cyan, sky blue 3, floral white, and turquoise modules. E) Scatter plots depicting uc.110 association with top positive correlation modules. Caption lists the module name, number of genes in the module, and the significance of uc.110 association with the module.

uc.110 has oncogenic effects in GBM.

To determine if uc.110 is an oncogene in GBM, we first used qPCR to investigate the expression of uc.110 using RNA in our banked tumor samples compared to normal brain cortex and cell lines compared to normal human astrocytes. We independently confirmed the results from our TCGA analysis showing uc.110 is highly upregulated in GBM tumors (Figure 4A and 4B). We then designed two siRNAs that target separate regions on the uc.110 RNA, one that begins at nucleotide 96/243 (si-uc.110-1) and one that begins at nucleotide 195/243 (si-uc.110-2), as well as a scrambled control (si-SCR) (Supplementary Figure 6A). We generated stable A172 and U251 GBM cell lines that express uc.110 (LV-uc.110) or the empty expression vector (LV-pCDH). We subjected these cell lines to siRNA transfection and assessed the effects on cell accumulation and invasion (Supplementary Figure 6B). We used qPCR to show that uc.110 is generally, though not uniformly, upregulated in GBM cells (Supplementary Figure 6). Based on these data, we prioritized the use of cell lines that overexpress uc.110 (A172, U251) for knockdown experiments, and cells that express low levels of uc.110 (U87, GSC-28) for overexpression experiments. We confirmed that siRNAs targeting of uc.110 leads to knockdown of uc.110 expression in A172 and U251 cells. (Figure 4C) We also confirmed that LV-uc.110 overexpresses uc.110, and that this overexpression rescues uc.110 bioavailability in A172 and U251 cells (Figure 4C).

Next, we performed cell counting assays [20, 37-39] to determine the effects of uc.110 knockdown and rescue on cell accumulation. When we reduced uc.110 expression, we reduced cell accumulation in A172 and U251 cells (Figure 4D). When we rescue uc.110 bioavailability, the cell accumulation phenotype is restored in A172 and U251 cells (Figure 4E). We then used AlamarBlue [40, 41] to measure cell viability. When we reduced uc.110 expression, A172 and U251 cell viability is reduced. We were able to rescue this phenotype by increasing uc.110 bioavailability (Figure 4F). We observed a similar phenotype in a glioma stem cell line that overexpresses uc.110 (GSC-34, Figure 4G).

We then investigated the invasive potential of uc.110 using a transwell invasion assay. [42-44] Knockdown of uc.110 reduced A172 and U251 cell invasion through a collagen IV matrix (Figure 4H). When uc.110 bioavailability was increased, a partial recovery of the phenotype is observed (Supplementary Figure 6F). Lastly, we overexpressed uc.110 in U87 and GSC-28 cells (Figure 4I) and determined that this leads to increased cell accumulation compared to the empty vector

after 7 days (Figure 4J) in U87 and GSC-28 cells. These data show that uc.110 has oncogenic effects in GBM cells.

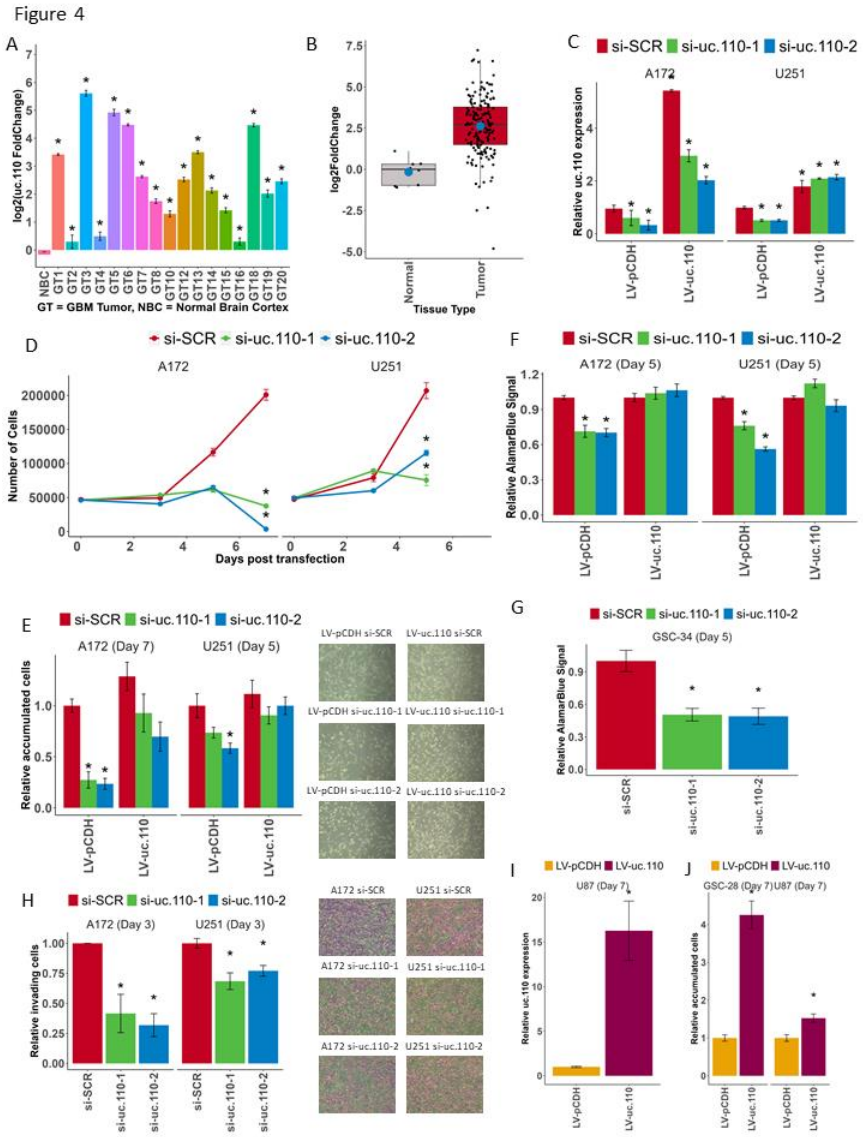


*Figure 4. Figure 4. The uc.110 TUCR operates as an oncogene. A) Bar graph depicting uc.110 upregulation in banked UVA GBM tumors versus normal brain cortex. B) Boxplot representing uc.110 expression in pooled tumors versus normal brain. Red boxes indicate an upregulated TUCR. C) Bar graph depicting qPCR validation of uc.110 siRNA knockdown and rescue in A172 and U251 cell lines. Facets represent cell lines. si-SCR = scrambled control siRNA (red), si-uc.110-1 = siRNA targeting uc.110 at nucleotide 96/243 (green), si-uc.110-2 = siRNA targeting uc.110 at nucleotide 195/243 (blue). D) Line graph showing that knockdown of uc.110 reduces A172 and U251 cell accumulation over a 5-7 day period. Facets represent cell lines. E) Bar graph depicting that the cell accumulation phenotype is rescued when uc.110 is overexpressed in the presence of siRNA. Facets represent cell lines. Images are representative of the listed sample. F) Bar graph showing that knockdown of uc.110 reduces A172 and U251 cell viability via Alamar Blue and can be rescued with uc.110 overexpression. Facets represent cell lines. G) Bar graph showing that knockdown of uc.110 reduces GSC-34 glioma stem cell viability via AlamarBlue. H) Bar graph showing knockdown of uc.110 reduces A172 and U251 cell invasion and migration. Images are representative of the listed sample. I) Bar graph showing the qPCR validation of overexpression in U87 cells. J) Bar graph showing that overexpression of uc.110 increases cell accumulation in U87 and GSC-28 cells. Facets represent cell lines. * = p < 0.05*

After determining that uc.110 displays an oncogenic phenotype *in vitro*, we sought to determine whether this effect is recapitulated *in vivo*. U251 GBM cells were transfected with si-uc.110-1 or si-uc.110-2. After 2 days, these cells were implanted into mice using intracranial injection (Supplementary Figure 7).[37, 38, 45, 46] Tumor growth was monitored by MRI and mouse survival was observed over a period of 70 days. Mice that were xenografted with U251 cells that were transfected with si-uc.110-1 and si-uc.110-2 expression developed smaller tumors, as depicted and quantified by MRI (Figure 5A, 5B). The mice that received si-uc.110 also displayed better overall survival than mice that received scrambled control siRNA cells (Figure 5C).

Figure 5

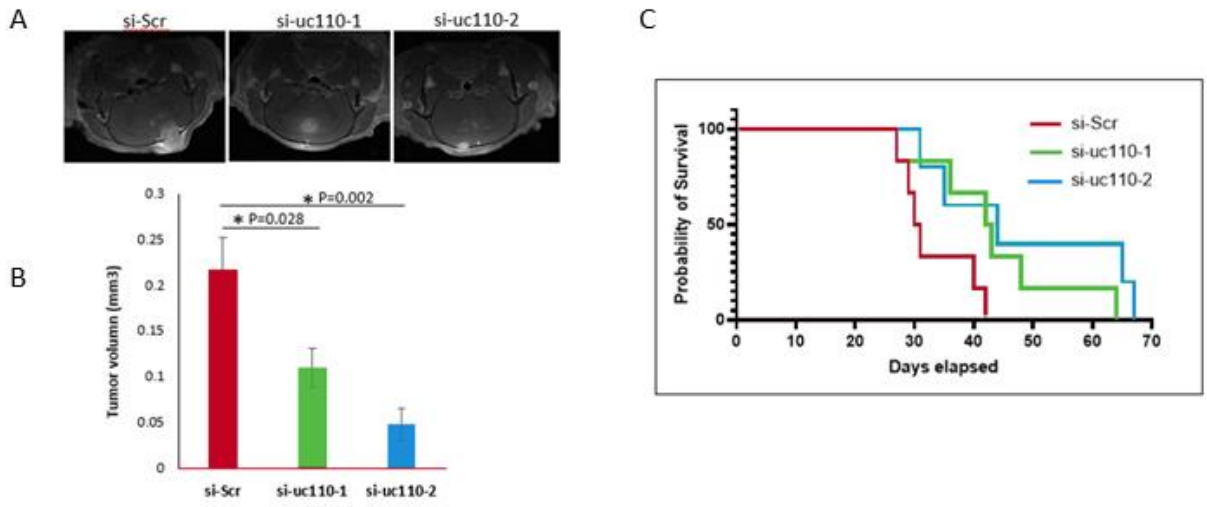


Figure 5. The uc.110 TUCR promotes tumor growth in vivo A) MRI images reveal a reduction in tumor size when uc.110 is knocked down via siRNAs. B) Bar graph showing that knockdown of uc.110 leads to a reduction in tumor volume. si-SCR = scrambled control siRNA (red), si-uc.110-1 = siRNA targeting uc.110 at nucleotide 96/243 (green), si-uc.110-2 = siRNA targeting uc.110 at nucleotide 195/243 (blue). C) Kaplan-Meier plot showing that xenograft knockdown of uc.110 leads to increased mouse survival.

uc.110 regulates the expression of the Wnt pathway member, Membrane Frizzled Related Protein (MFRP).

LncRNAs can have various functions that depend on their subcellular localization. Nuclear lncRNAs are usually involved in transcriptional regulation, while cytosolic lncRNAs are usually involved in translational and spatial regulation. [2] We fractioned four GBM cell lines (A172, U251, U87, U1242) in nuclear and cytosolic fractions. When compared to nuclear (U44, U48) and cytosolic (GADPH, PPIA) controls, uc.110 appears to be localized to both the nucleus (mainly in U87, U251, and U1242 cells), and the cytoplasm (mainly in A172 cells) (Supplementary Figure 6G). We then performed RNA-Seq on A172 cells that had been transfected with si-SCR, si-

uc.110-1, or si-uc.110-2 for 48 hrs. and found several genes that are deregulated when uc.110 expression is downregulated (Figure 6A). To identify genes that are particularly related to uc.110 function, we focused on genes that demonstrated coregulation with uc.110 in our WGCNA analysis (Figure 3E). Of particular interest was the membrane frizzled related protein, also known as MFRP. [47, 48] MFRP serves as a shuttle for the Wnt-ligand, and functions as an activator of the Wnt-signaling pathway. This gene was the only gene in our analysis that correlated with uc.110 expression, was upregulated in GBM Tumors, and downregulated when uc.110 is knocked down in A172 cells. (Figure 6B).

uc.110 sponges the tumor suppressor microRNA miR-544 to increase the bioavailability of MFRP and WNT activity in GBM.

One common lncRNA mechanism of action is as a miRNA sponge, acting as a binding competitor for various miRNAs and therefore increasing the bioavailability of those miRNAs' targets.[2, 49-52] We hypothesized that uc.110 may function in a similar manner by sponging miRNAs away from MFRP transcripts, as their expression relationship is consistent with such an interaction. We hypothesized that a tumor suppressor miRNA is able to successfully target and suppress MFRP in the normal brain (Supplementary Figure 8A). This leads to downstream activation of Wnt target genes involved in biological processes such as cell accumulation, invasion, and stem cell differentiation (Supplementary Figure 8B). We further hypothesized that in glioma tumors, uc.110 is activated and acts as a binding competitor for this miRNA (Supplementary Figure 8C), increasing the bioavailability of MFRP and increasing Wnt pathway signaling (Supplementary Figure 8D). To test these hypotheses, we examined the presence of putative miRNA binding sites for tumor suppressor miRNAs within the transcripts of uc.110 and MFRP. We found that miR-544 has putative binding sites in uc.110 and is predicted to target MFRP (Supplementary Figure 9A). We first investigated the functional effects of miR-544 in GBM cells. Transfection of miR-544 into U251, A172, and T98G GBM cell lines reduced cell accumulation after 5 days (Figure 6C). Expression of both uc.110 and MFRP was reduced when transfected with miR-544 or si-uc.110 (Figure 6D). To further test the hypotheses, we asked if miR-544 targets both uc.110 and MFRP, and if this binding affects Wnt signaling. We first measured target binding via transfected Renilla luciferase constructs containing either the uc.110 (Figure 6E) or MFRP (Figure 6F) binding sites. We then mutated the binding sites for MFRP (Supplementary Figure 9B) and uc.110 (Supplementary Figure 9C) and assessed signal strength again. Our data show that miR-544 binds both uc.110 and MFRP, and that this binding is lost when the miRNA binding sites are mutated. (Figure 6G).

Figure 6

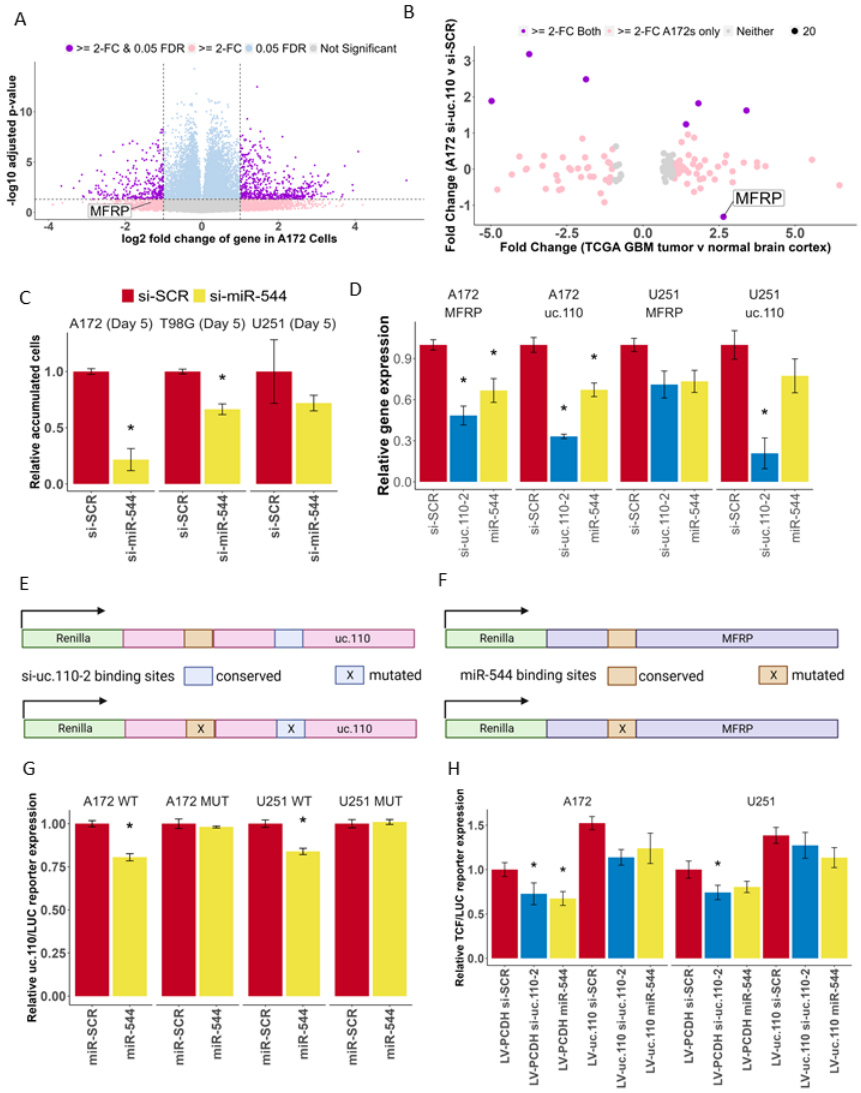


*Figure 6. The uc.110 TUCR activates Wnt-signaling by sponging miR-544 from membrane frizzled related protein (MFRP) 3'UTR. A) Volcano plot depicting transcriptome deregulation in RNA-Seq data on A172 cells transfected with si-uc.110. Purple dots represent genes that are significantly deregulated >= 2-fold. Blue dots represent genes that are significantly deregulated. Pink dots represent genes that are deregulated >= 2-fold. Gray dots are neither deregulated nor significant. B) Dot plot showing that, of the genes that predicted miR-544 targets, MFRP is the only gene that is upregulated in GBM Tumors and downregulated when uc.110 is downregulated. Purple dots represent predicted miR-544 targets that are deregulated in A172 cells from 6A and TCGA RNA-Seq data. Pink dots represent predicted miR-544 targets that are deregulated in A172s from 6A only. C) Bar graph showing that miR-544 transfection reduces cell accumulation in A172, T98G, and U251 cells, confirming its tumor suppressor role. Facets represent cell lines. si-SCR = scrambled control siRNA (red), si-uc.110-2 = siRNA targeting uc.110 at nucleotide 195/243 (blue). miR-544 = miR-544 (yellow). D) Bar graph showing that transfection with miR-544 or si-uc.110-2 reduces uc.110 and MFRP expression. Facets represent genes and cell lines. E) Schematic depicting the uc.110 luciferase construct used to demonstrate binding to miR-544 and si-uc.110-2. Binding of miR-544 to binding sites (orange) leads to a degradation of construct and a reduction in Renilla signal (Green) F) Schematic depicting the MFRP luciferase construct used to demonstrate binding miR-544 G) Bar graph showing that transfection of miR-544 reduces uc.110 and MFRP luciferase signal, and that mutating miR-544 binding sites rescues luciferase signal. Facets represents cell lines and miR-544 binding site mutation status. H) Bar graph showing that transfection of si-uc.110-2 and miR-544 reduces TCF/LEF reporter signal. Signal is rescued when uc.110 is overexpressed in the presence of siRNA or miR-544. * = p< 0.05*

Lastly, we asked if uc.110 expression alters Wnt pathway activity. To answer this question, we studied one of the most established downstream targets of Wnt-signaling, the T cell factor/lymphoid enhancer factor family (TCF/LEF). When Wnt-signaling is activated, TCF/LEF is produced downstream and activates Wnt-signaling target genes. Therefore, TCF/LEF activity can be used as a proxy for pathway activity and can be measured with a TCF/LEF luciferase reporter assay. (Supplementary Figure 9D). The activity of this reporter can be regulated by either directly reducing Wnt bioavailability with miR-544 or indirectly by targeting uc.110 with siRNA. If upstream Wnt signaling is reduced, the luciferase construct will bind fewer activators and exhibit decreased signal (Supplementary Figure 8D). Likewise, we would expect that overexpression of uc.110 would rescue the bioavailability of MFRP and consequently also downstream activation of the TCF/LEF construct. We found that transfection of A172 and U251 cells with si-uc.110 and miR-544 reduced reporter activity, and that this effect can be rescued via uc.110 overexpression. (Figure 6H). These data taken in conjunction provide strong support for a miRNA sponge model for the uc.110 oncogene. Altogether, the above data demonstrate an important role for uc.110 in regulating the Wnt pathway in GBM by sponging the Wnt inhibitory miRNA miR-544.

DISCUSSION

This study investigated Transcribed Ultraconserved Regions (TUCRs), a set of putative long non-coding RNA sequences that are fully conserved across human, mouse, and rat genomes. These TUCRs are distinct due to their exceptional conservation, which often signifies functional importance. Despite their potential significance, TUCRs have been minimally explored, especially in relation to cancer. Of note, the findings of this study represent the first of their kind on TUCRs in gliomas. They contribute critical new insights into an uncharted area of glioma biology, while also providing a novel framework for studying TUCRs in other cancers and other diseases, where they are also understudied.

We found that TUCRs are located across the genome, resistant to variation, and actively transcribed. We manually annotated each as exonic, intronic, exonic/intronic, or intergenic. We identified distinct signatures for intergenic and intragenic (exonic, intronic, exonic/intronic) RNAs, which suggests that they may have different biological functions. Intragenic TUCRs are expressed at a level that is most like coding genes, while the intergenic TUCRs more resemble lncRNAs. We then performed the first analysis of TUCR expression in gliomas and found that the

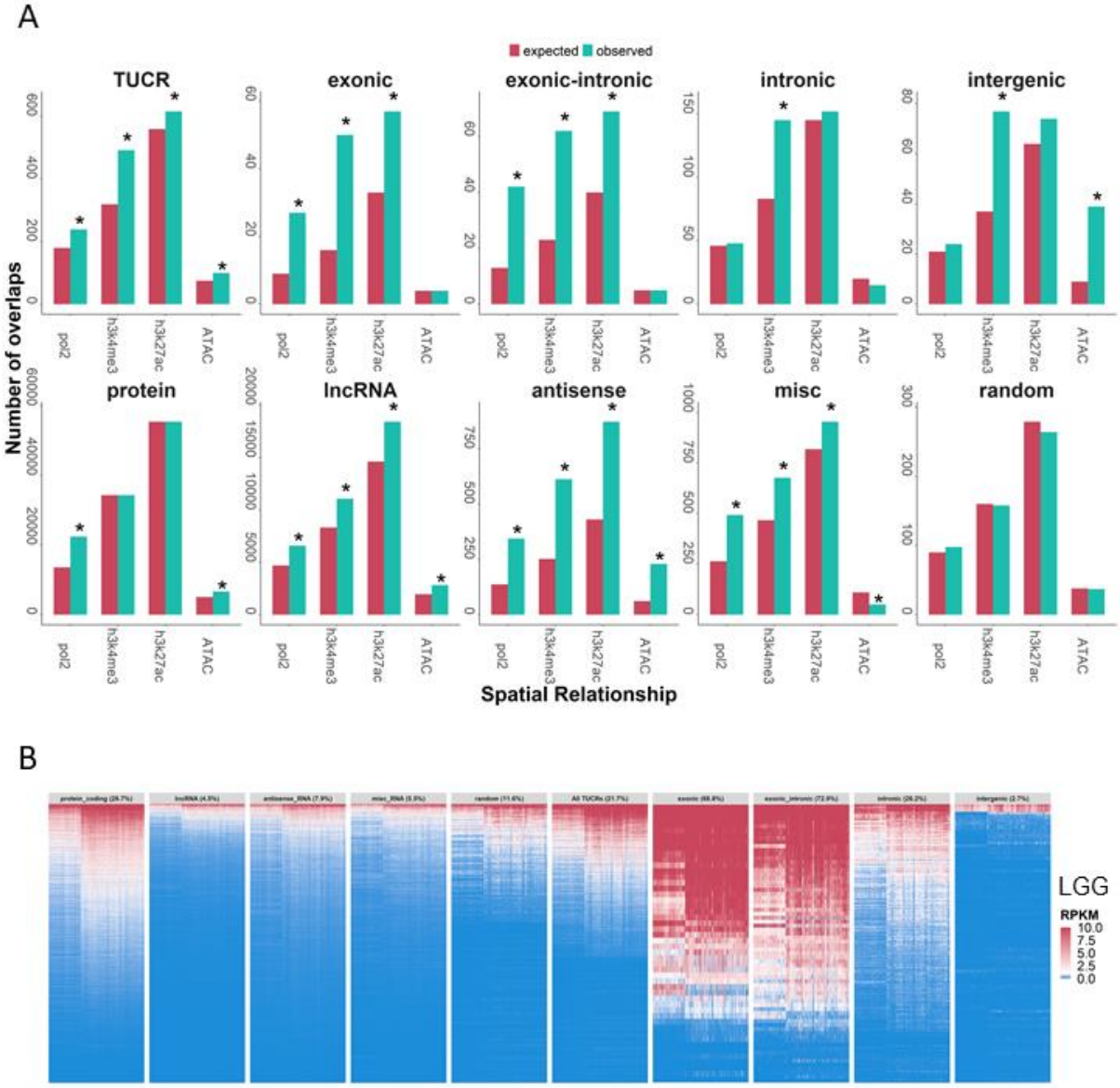
majority of TUCRs are deregulated >= 2-fold in GBM and LGG, with a 56% overlap. This shows that TUCRs are not only expressed, but also frequently deregulated in gliomas compared to normal brain tissue. This is critical, as their high degree of conservation and deregulation suggests that they may serve critical biological functions. We then extended our analysis to TUCR correlation with patient survival. In GBM, the extremely short survival times (15 months) limit the detection of significant correlations. However, patients with LGG live substantially longer (84 months), and therefore more TUCRs are associated with patient outcomes in this disease, suggesting a potential impact on glioma patients' prognoses and indicating possible novel biomarkers. Another facet of our research involved predicting the functions and mechanisms of action of TUCRs in gliomas. We studied this for the first time in gliomas WGCNA workflows to cluster TUCRs and provide functional predictions based on shared functions between coregulated genes. This approach identifies a wide range of potential functions for TUCRs, encompassing activities such as nucleic acid binding regulation, stem cell differentiation, organ development, immune response, and cell signaling.

We found intergenic TUCRs to be of notable interest because they resemble lncRNAs, but experience less sequence variation. Notably, these TUCRs do not overlap with known genes, suggesting they might represent novel lncRNAs. Of these TUCRs, uc.110 is the most upregulated in both GBM and LGG. Knocking down uc.110 reduces cancer cell characteristics *in vitro* and *in vivo* and improves survival in mouse models. On the other hand, increasing uc.110 expression increases malignancy in cells that do not express it, further indicating its potential oncogenic role. We explored uc.110's function via WGCNA, revealing its membership in modules associated with oncogenic nucleic acid binding. We integrated these data with transcriptome deregulation data (RNA-Seq) post-uc.110 knockdown, revealing a close relationship between uc.110 and the oncogenic membrane frizzled-related protein (MFRP). This protein is involved in activating the Wnt-signaling pathway, impacting cell proliferation, invasion, migration, and stem cell differentiation. From these data we hypothesized that uc.110 might sponge tumor suppressor miRNAs from MFRP, enhancing Wnt signaling activation. Accordingly, we demonstrated that one mechanism of action for the uc.110 oncogene is as a miRNA sponge for miR-544, therefore increasing the bioavailability of MFRP.

In conclusion, our results suggest that TUCRs are an important class of regulatory RNAs. They are more highly conserved than typical genes and more resistant to variation, which suggests biological importance. They are perturbed in gliomas, and this perturbation is associated with clinical outcomes. Our predicted functions reveal that TUCRs are widely involved in cancer-related biological processes. Also, some TUCRs previously thought to be intergenic may represent previously undiscovered genes. Our findings also identify and characterize uc.110 as a new oncogene in gliomas. Each of the experiments performed in our study represents the first of its kind in gliomas. We have developed and presented novel methods for studying TUCRs that can be used in other cancers and other diseases, where TUCRs remain very understudied. These methods and the data derived from them represent a “TUCR database” that will serve the scientific community in future TUCR studies in gliomas and other diseases, where they remain unstudied or understudied.

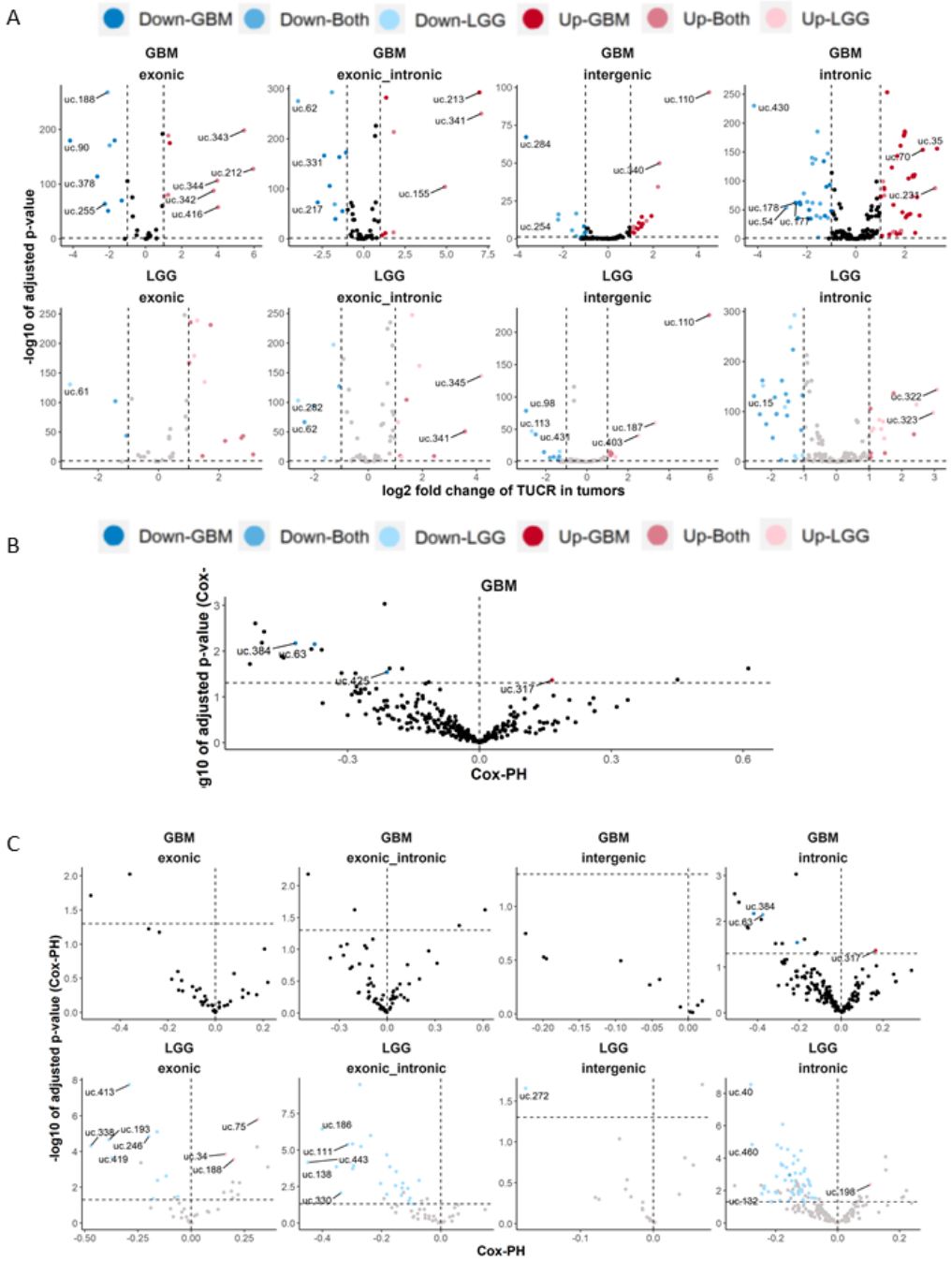
SUPPLEMENTARY DATA

Supplementary Figure 1



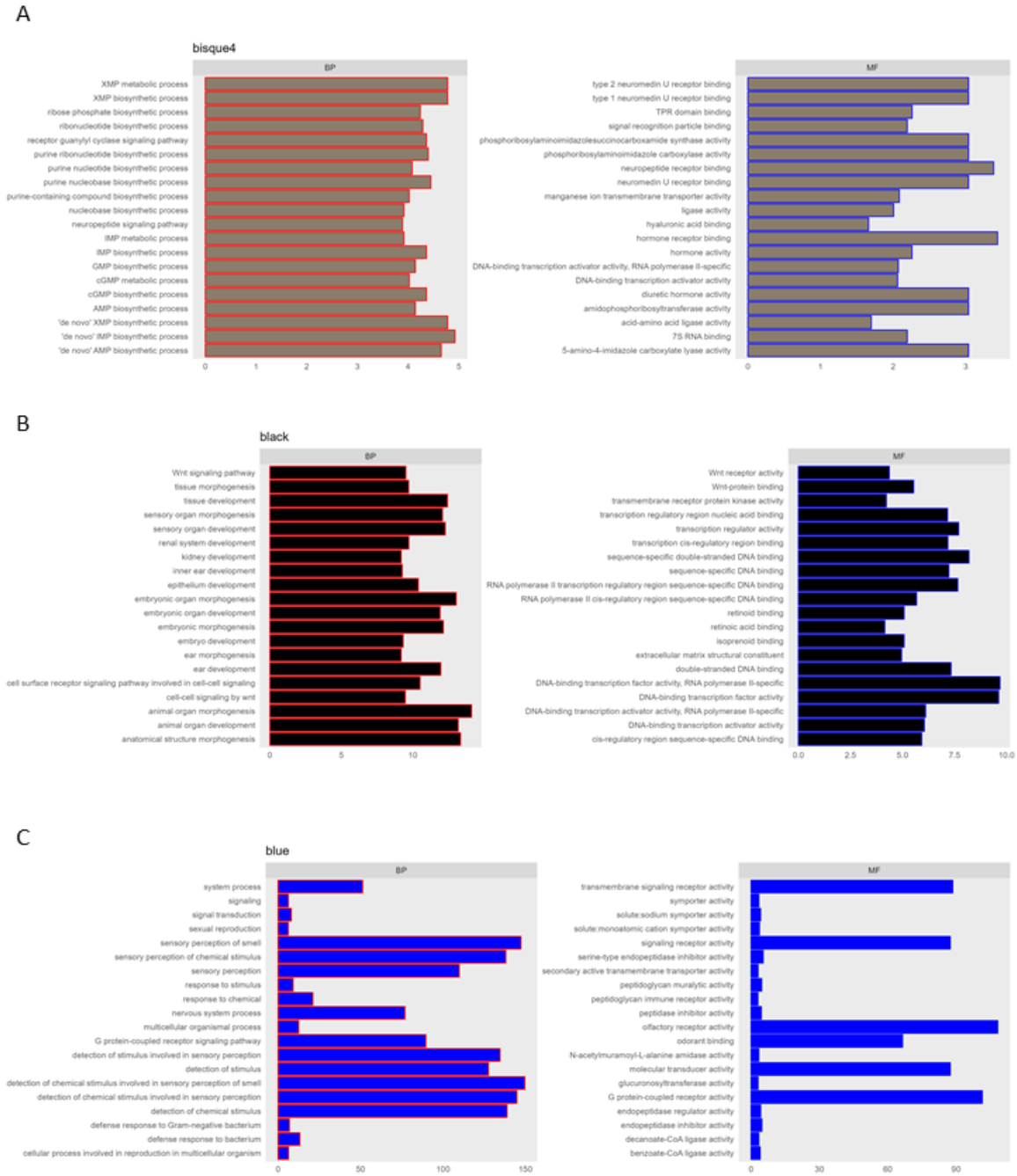
*Supplementary Figure 1. Annotation, localization and expression of TUCRs in GBM and LGG (cont.). A) Bar chart showing that TUCRs are enriched for markers for open and active chromatin in GBMU87 cells, suggesting that they represent transcriptionally active sites. Red bars represent chi-square expected overlaps, and teal bars represent observed values. B) Heatmap representing TUCR absolute expression (RPKM) across multiple gene annotations. Blue represents poorly expressed genes (<1 RPKM), White/Pink genes are moderately expressed (>=1 RPKM) and Red represents highly expressed genes (RPKM >=10). TUCRs demonstrate an expression profile that is comparable with protein coding genes. * = p < 0.05*

Supplementary Figure 2



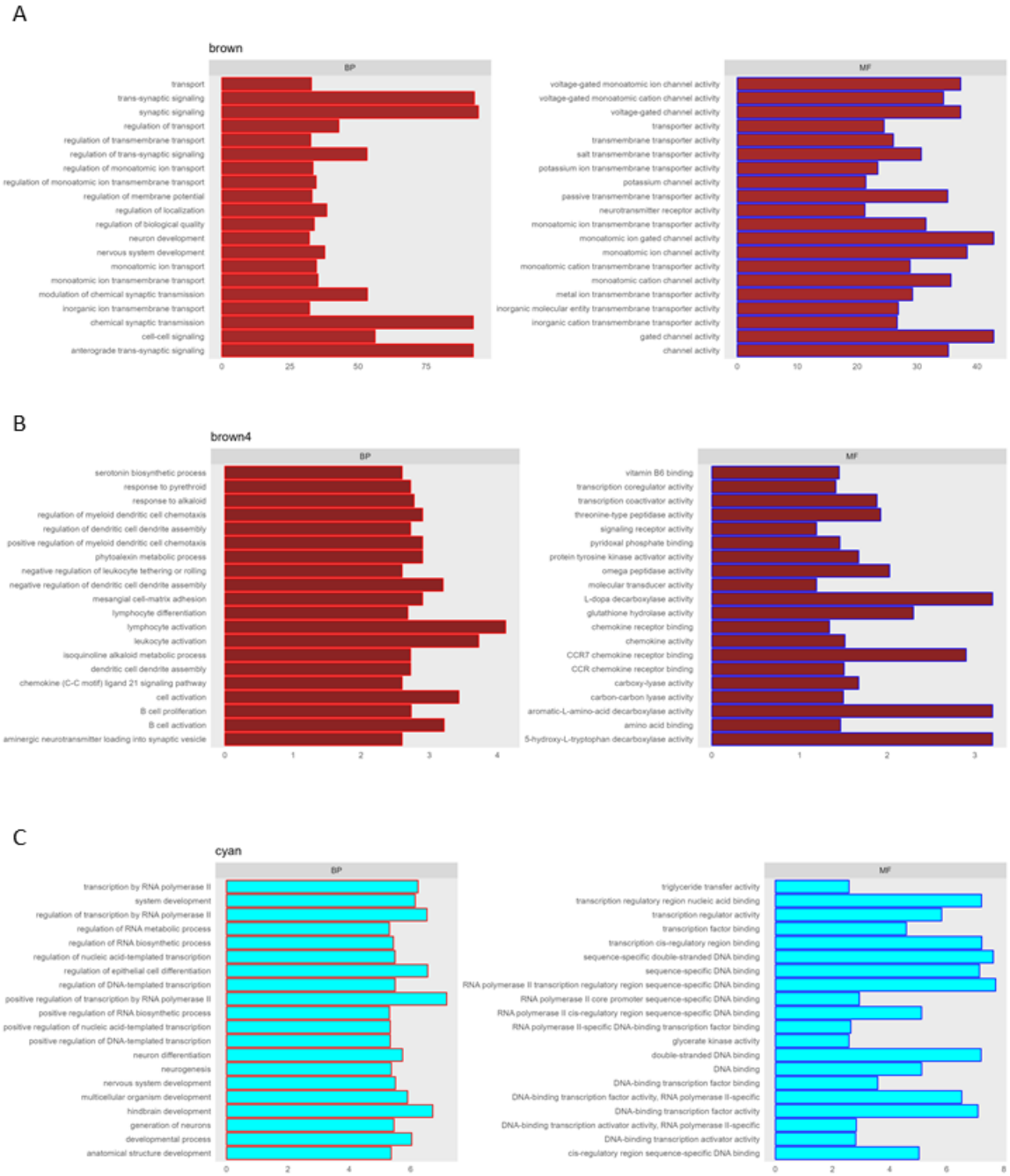
Supplementary Figure 2. TUCRs are deregulated and associated with patient outcomes in gliomas and may have broad functional roles (cont.) All experiments were performed using TCGA GBM and LGG RNA-Seq data. A) Volcano plots showing that TUCRs are deregulated in every TUCR annotation category in GBM and LGG. Red dots are upregulated. Blue are downregulated. B) Volcano plot showing that few TUCRs are significantly associated with patient outcomes in GBM. Red dots represent TUCRs significantly associated with poor prognosis. Blue dots represent TUCRs significantly associated with good prognosis. C) Volcano plot showing that TUCRs in every TUCR annotation category are associated with survival in gliomas. Red dots represent TUCRs significantly associated with poor prognosis. Blue dots represent TUCRs significantly associated with good prognosis.

Supplementary Figure 3-1



Supplementary Figure 3-1. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

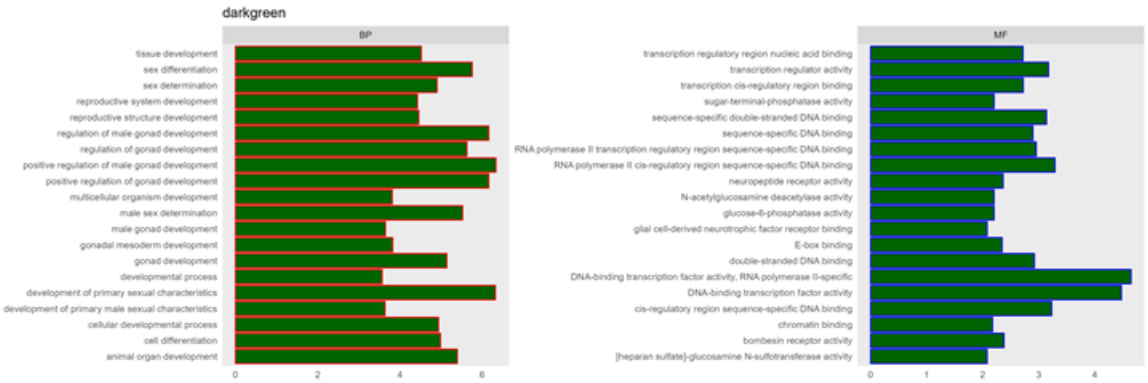
Supplementary Figure 3-2



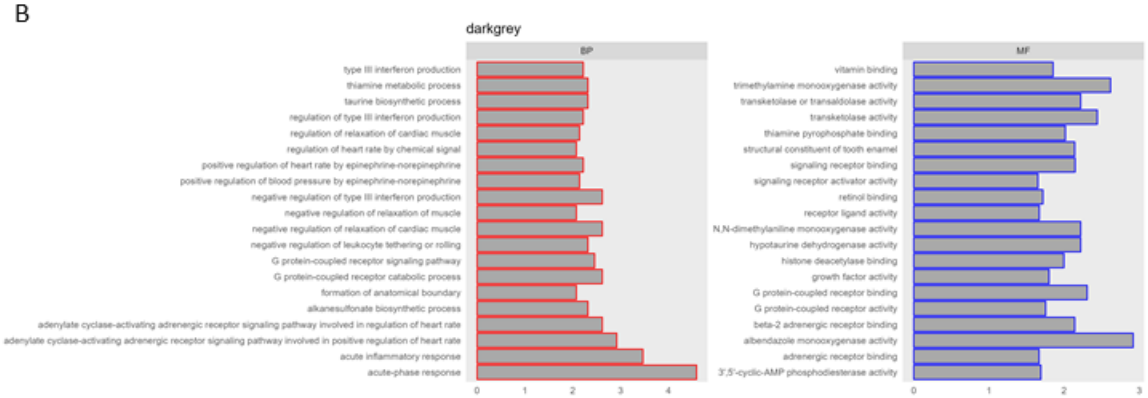
Supplementary Figure 4-2. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

Supplementary Figure 3-3

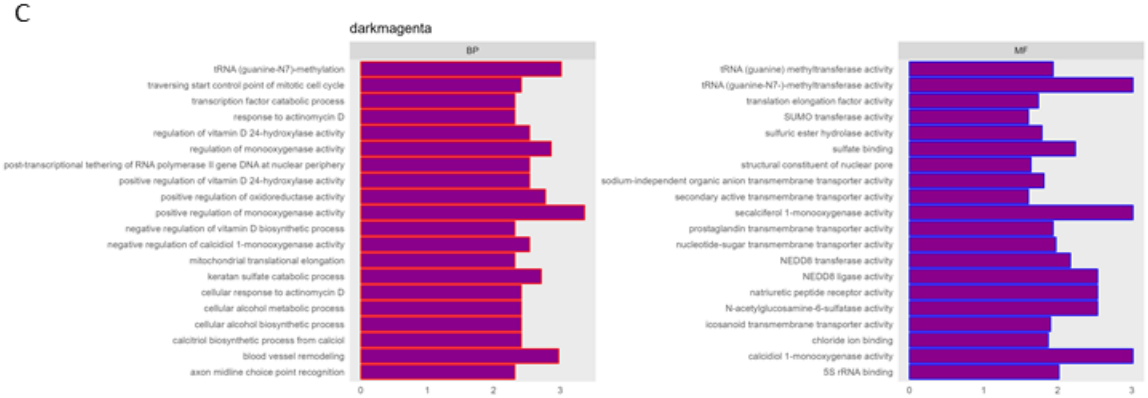
A



B

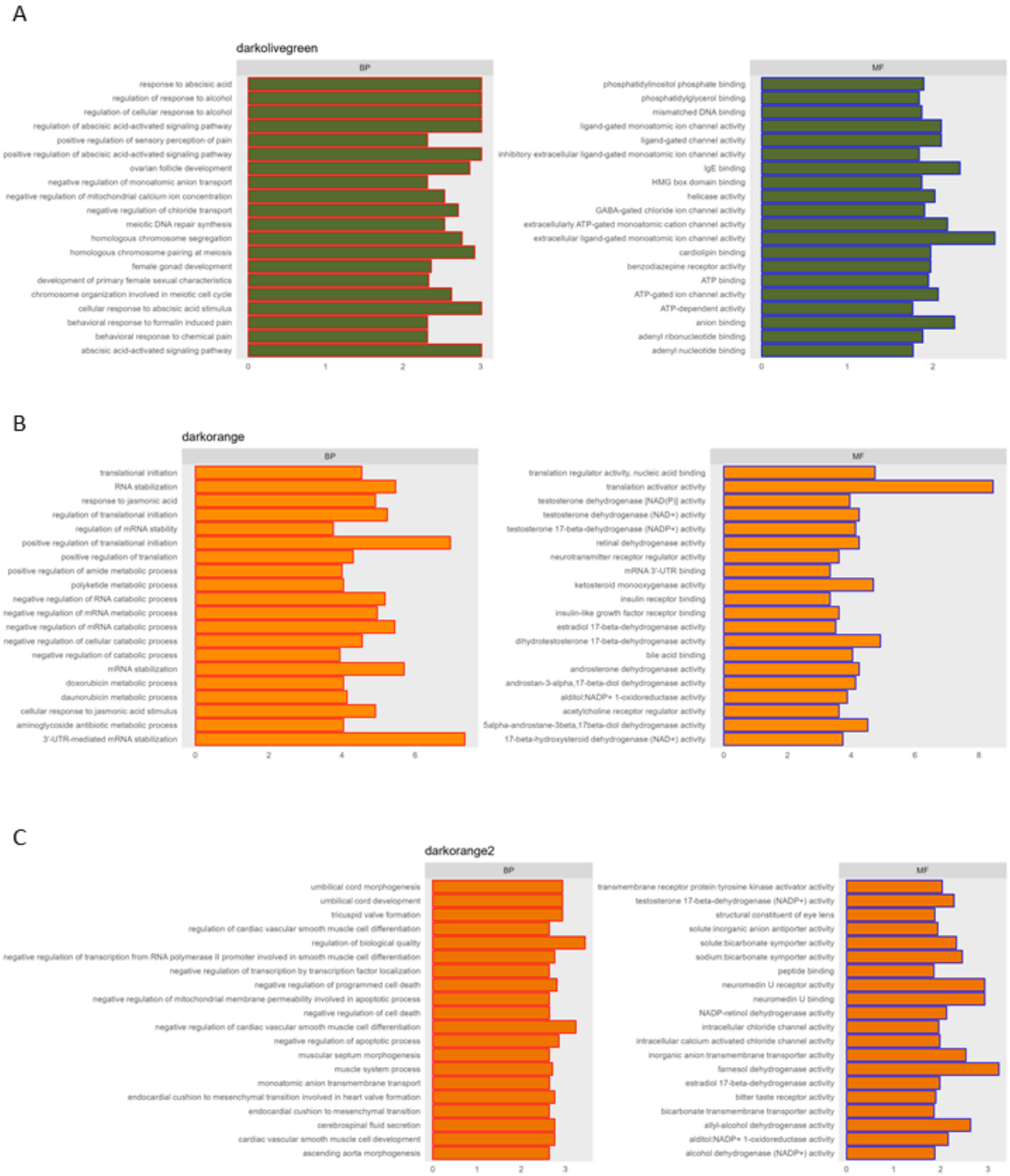


C



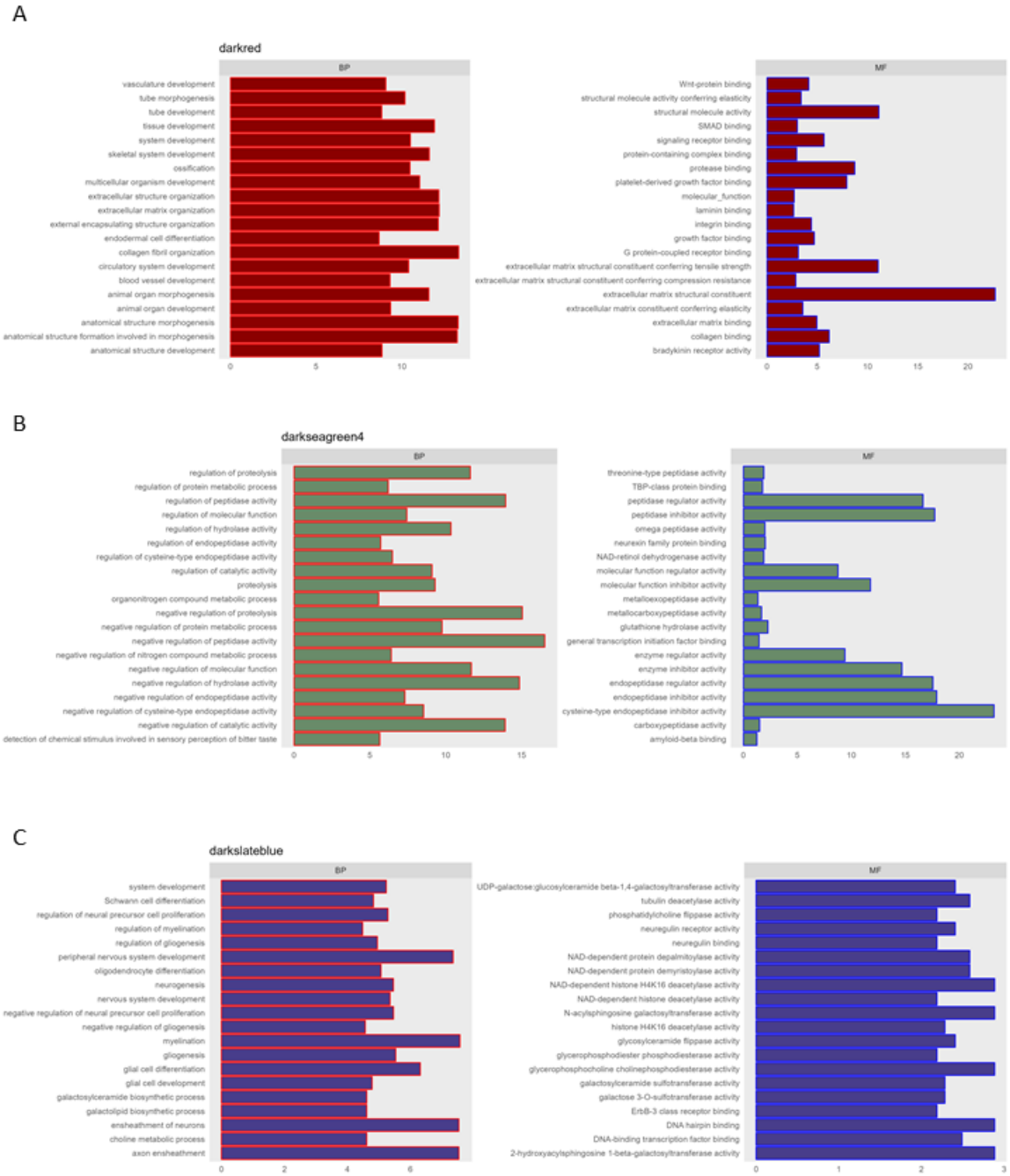
Supplementary Figure 5-3. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

Supplementary Figure 3-4



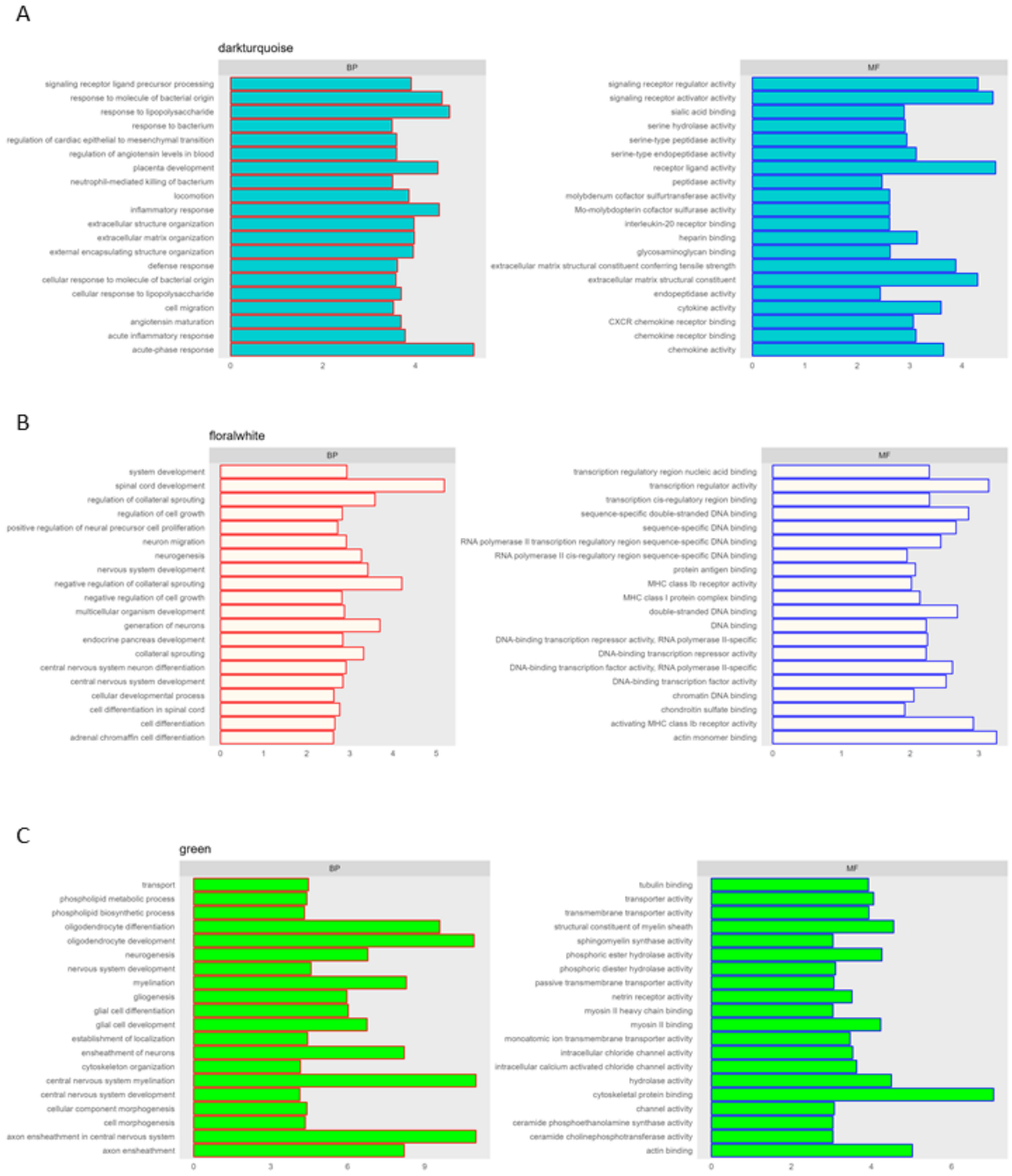
Supplementary Figure 6-4. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

Supplementary Figure 3-5



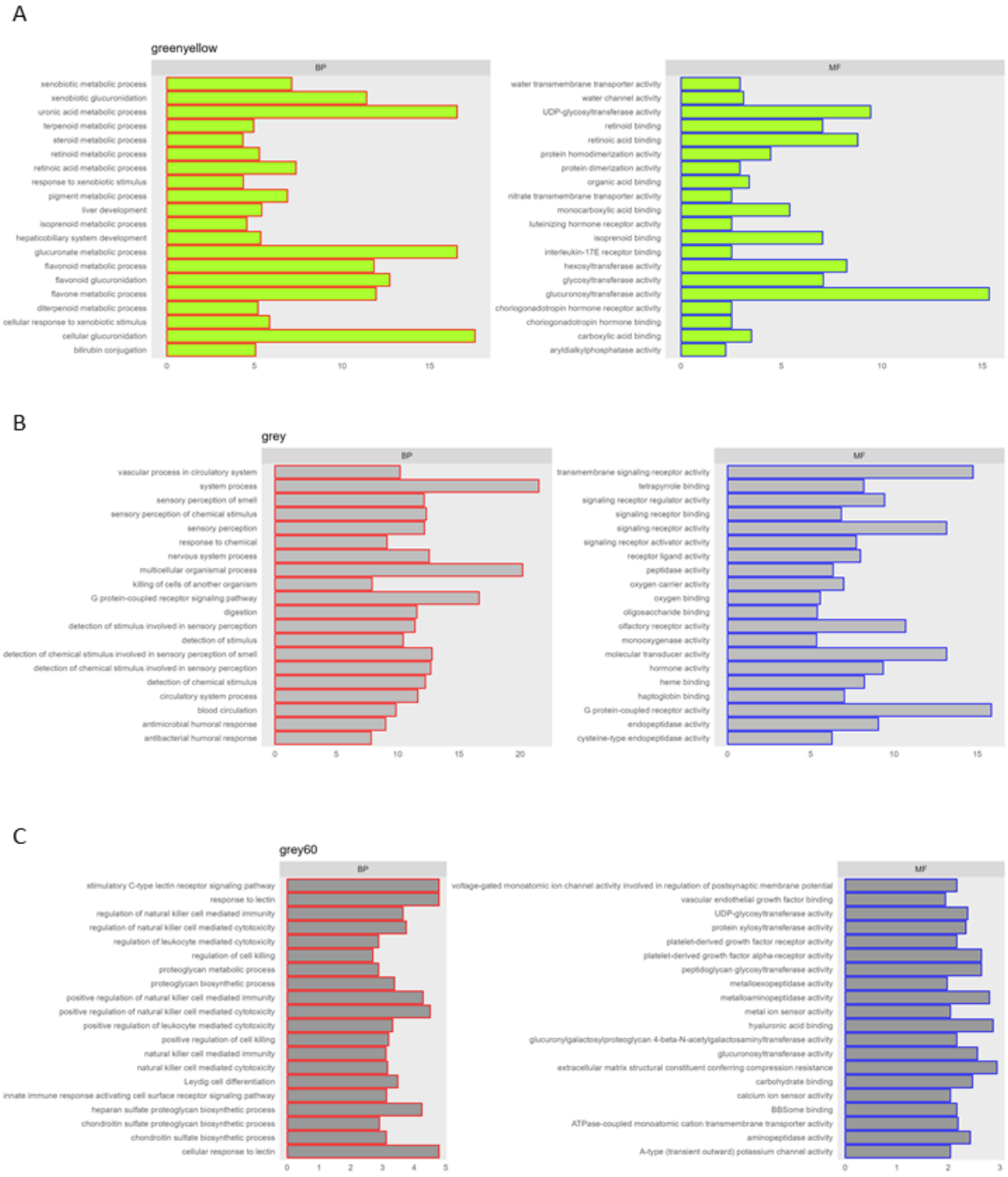
Supplementary Figure 7-5. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

Supplementary Figure 3-6



Supplementary Figure 8-6. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

Supplementary Figure 3-7

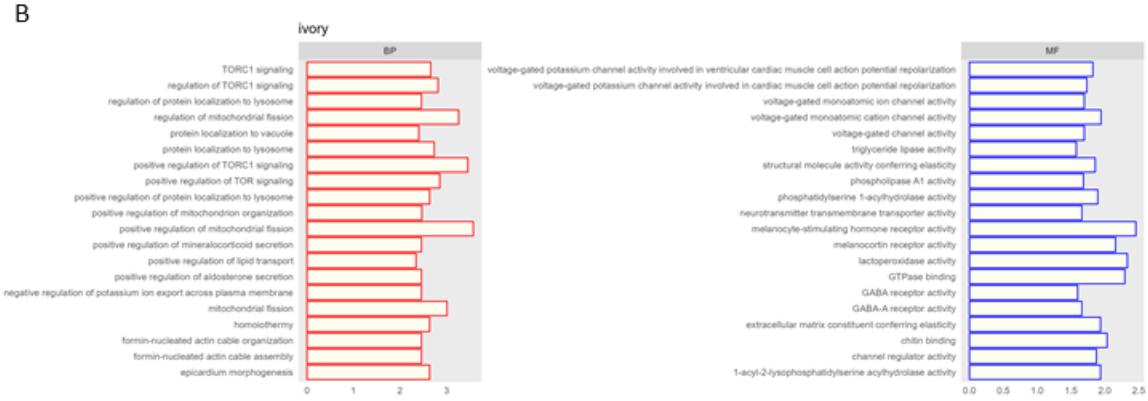


Supplementary Figure 9-7. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

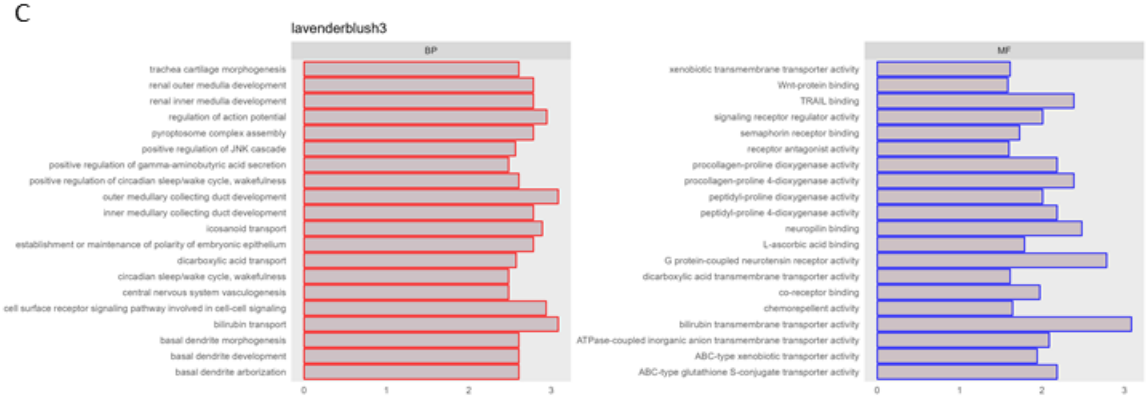
Supplementary Figure 3-8

A

B

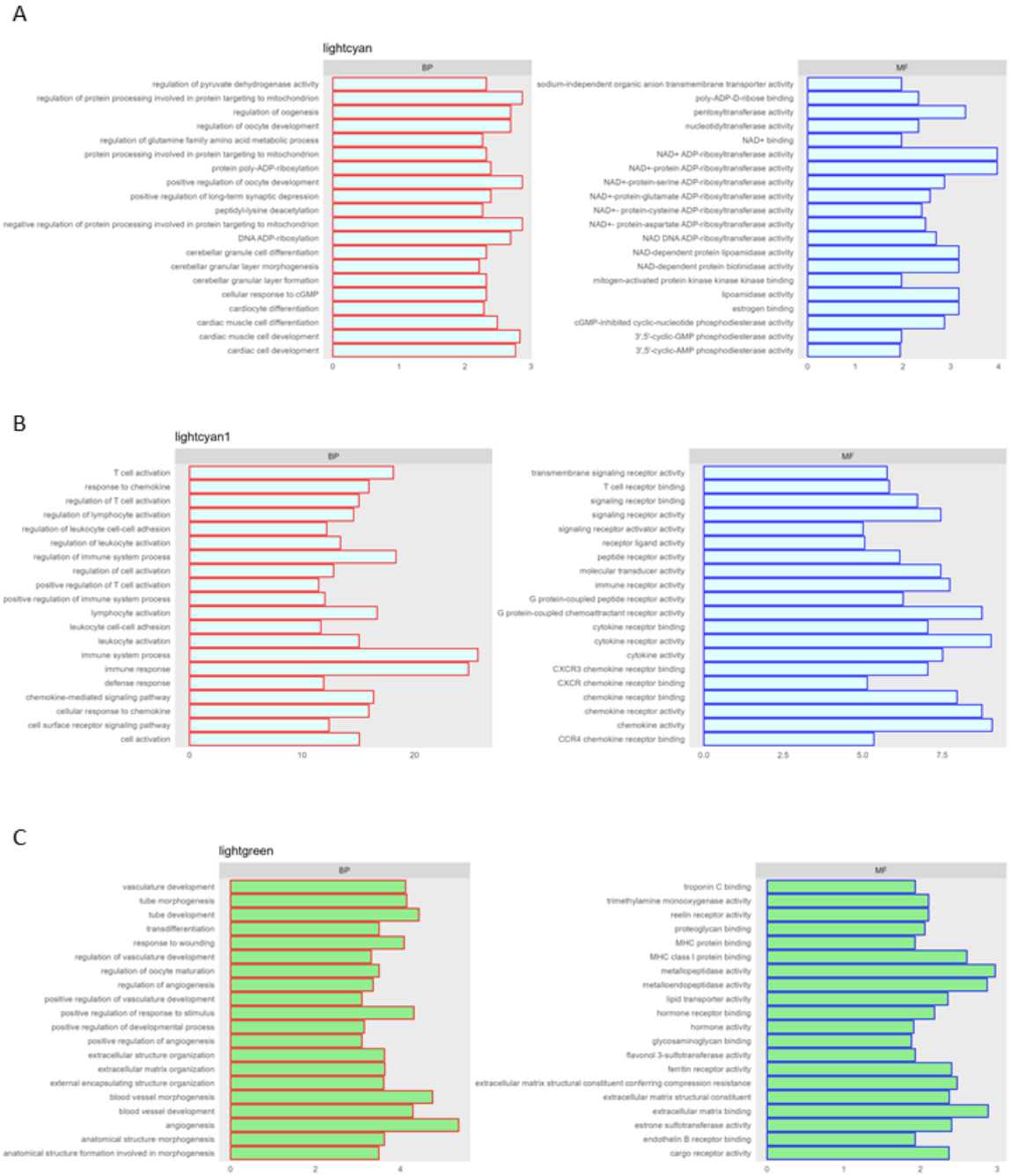


C



Supplementary Figure 10-8. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

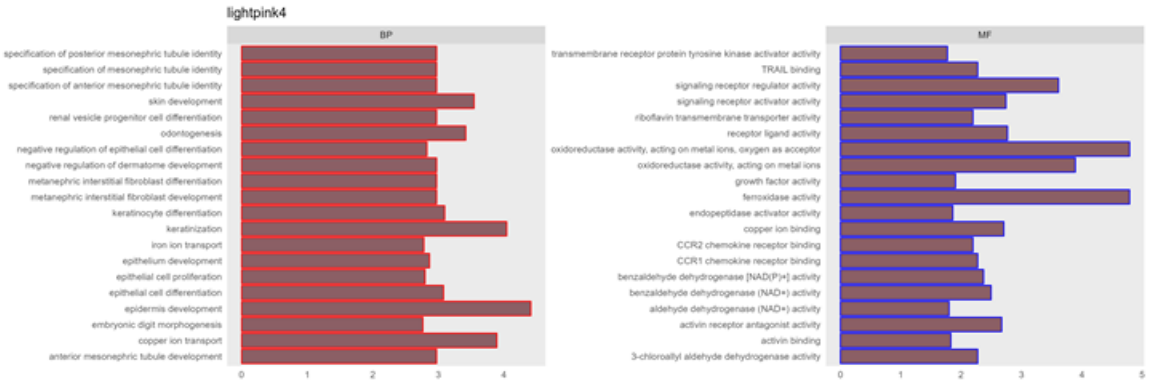
Supplementary Figure 3-9



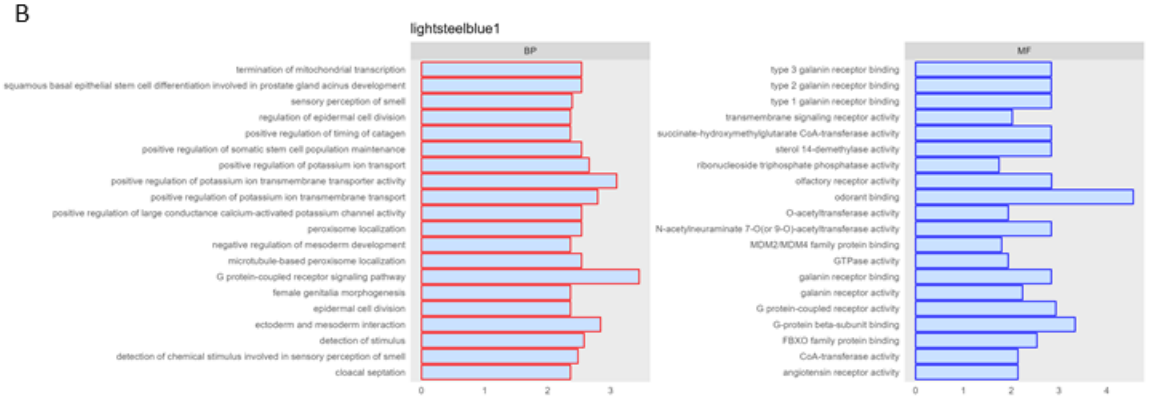
Supplementary Figure 11-9. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

Supplementary Figure 3-10

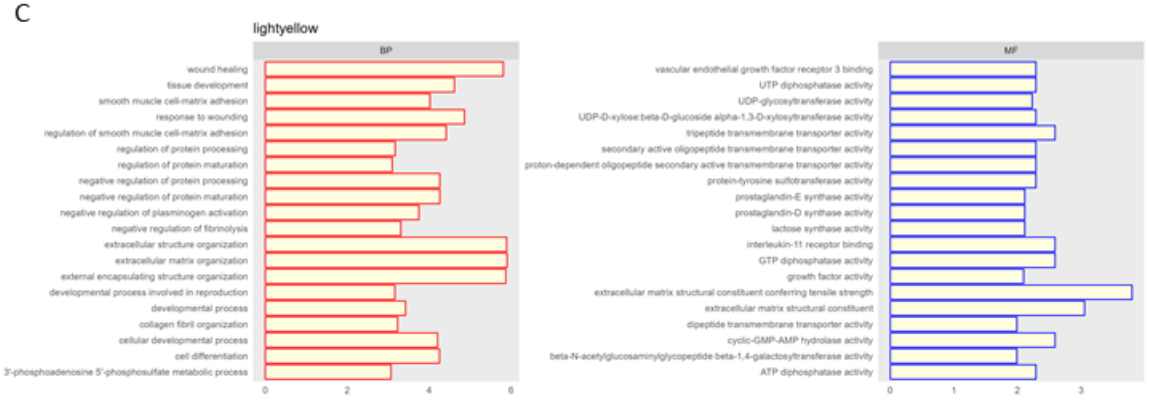
A



B



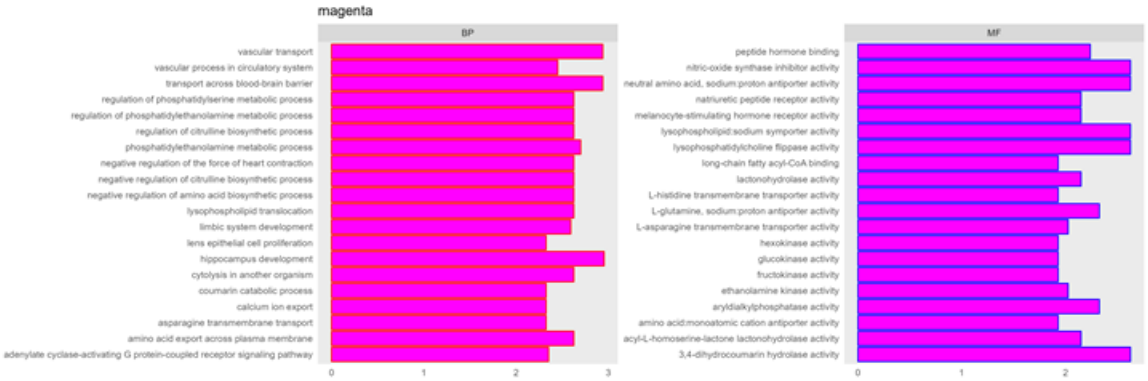
C



Supplementary Figure 12-10. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

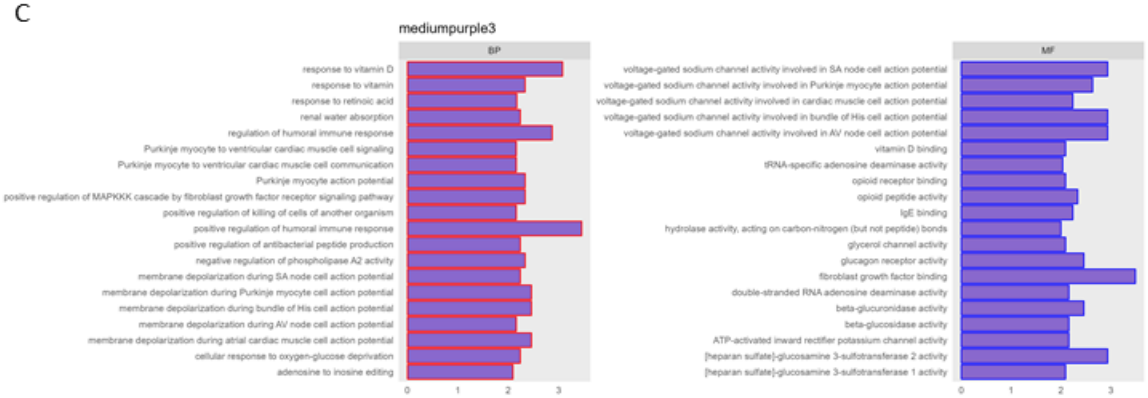
Supplementary Figure 3-11

A



B

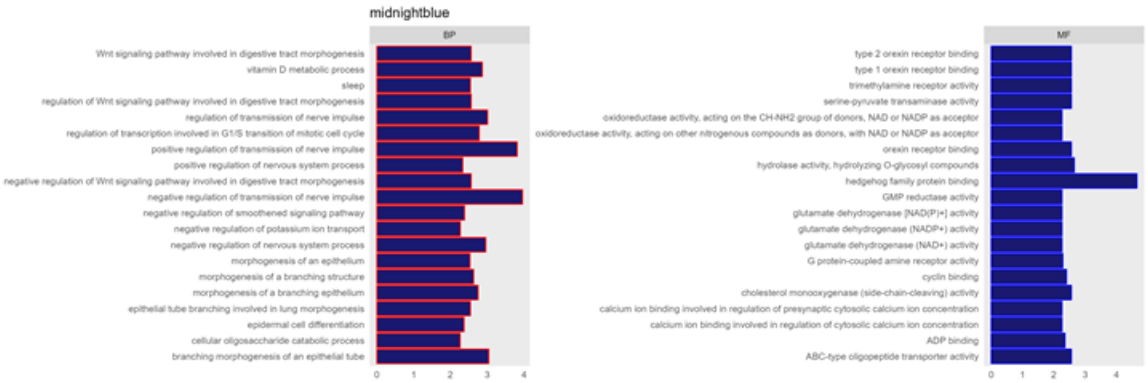
C



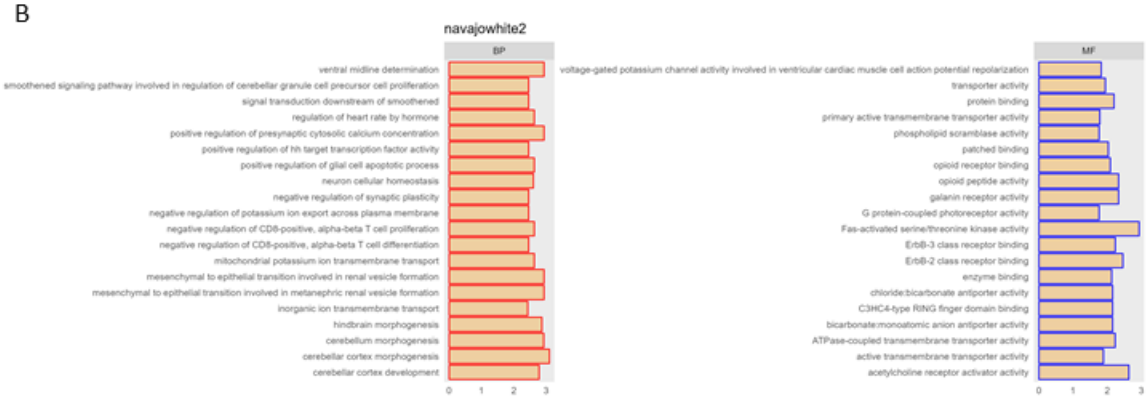
Supplementary Figure 13-11. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

Supplementary Figure 3-12

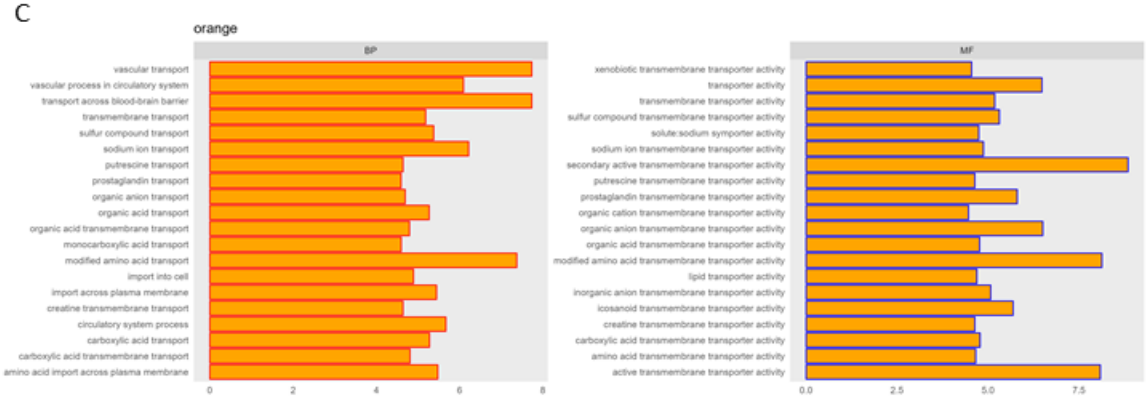
A



B

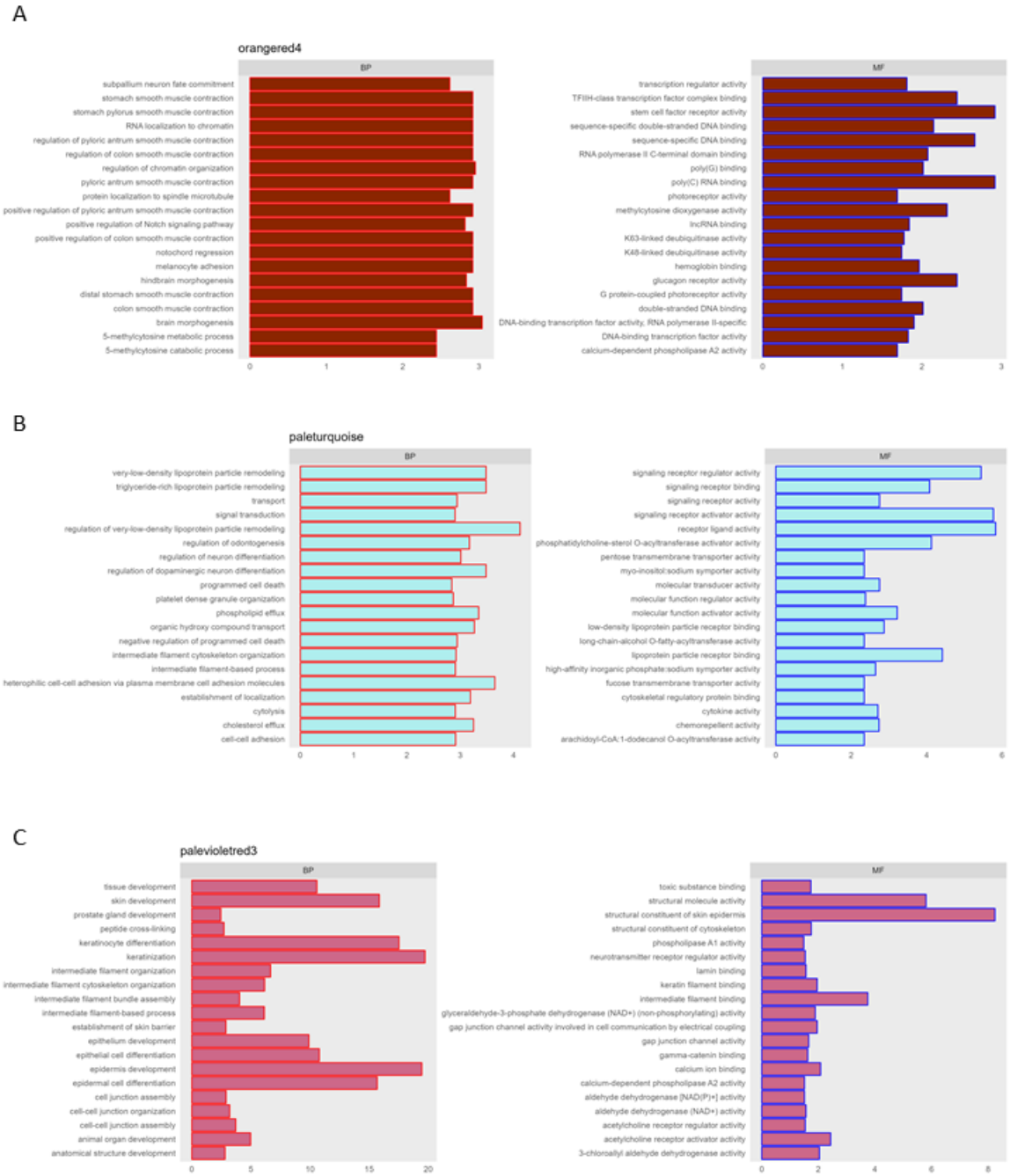


C



Supplementary Figure 14-12. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

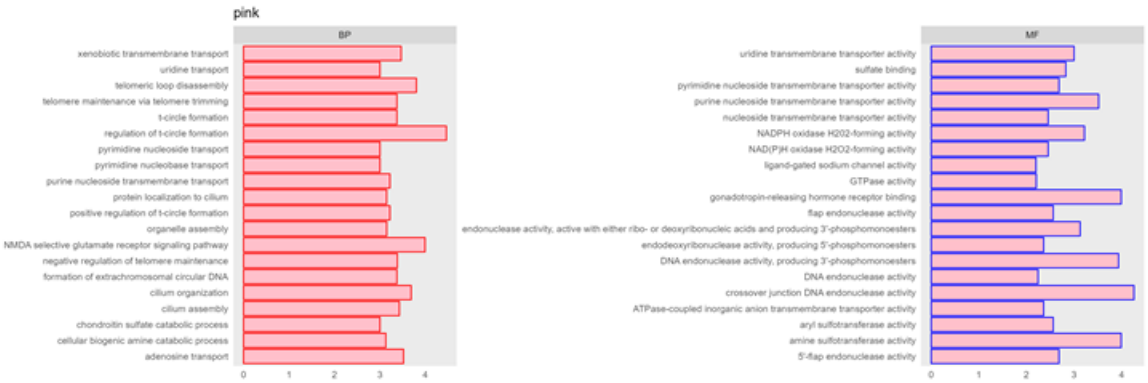
Supplementary Figure 3-13



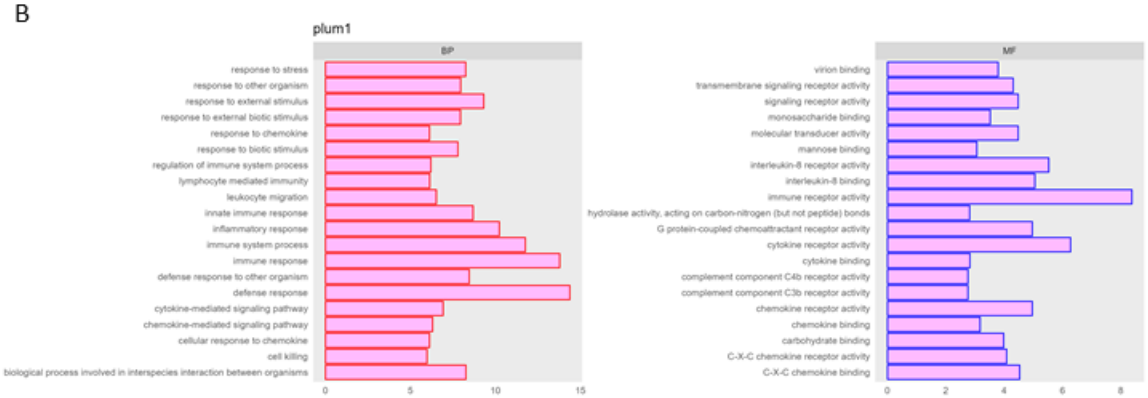
Supplementary Figure 15-13. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

Supplementary Figure 3-14

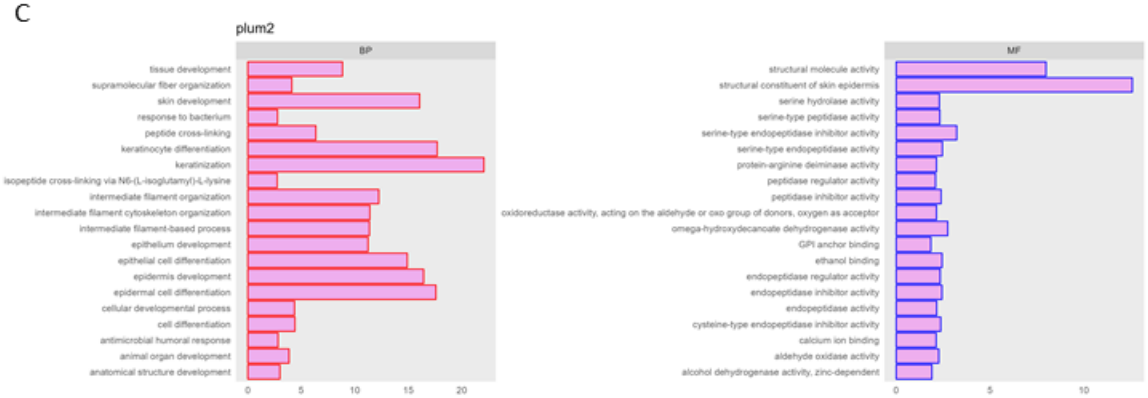
A



B



C

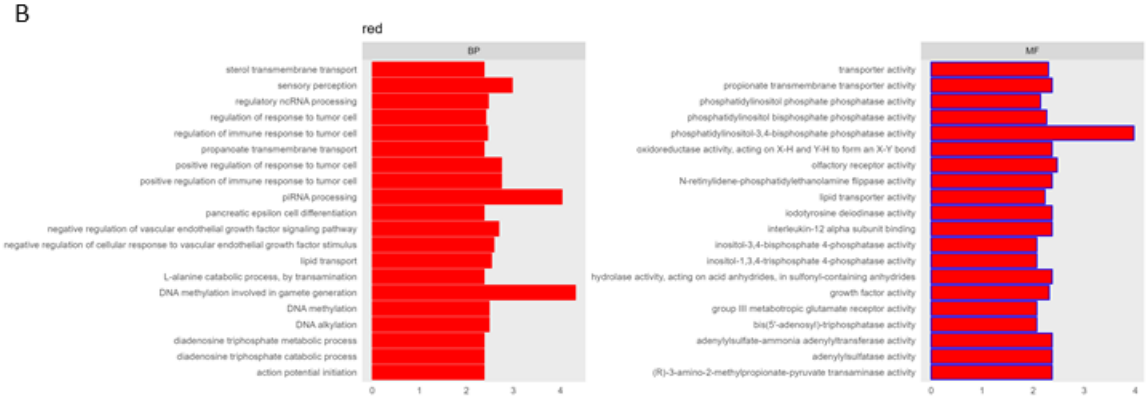


Supplementary Figure 16-14. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

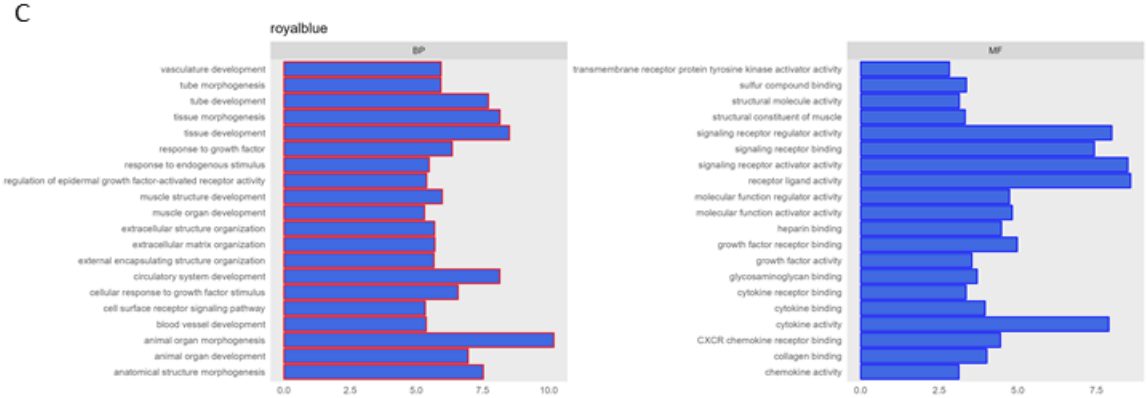
Supplementary Figure 3-15

A

B



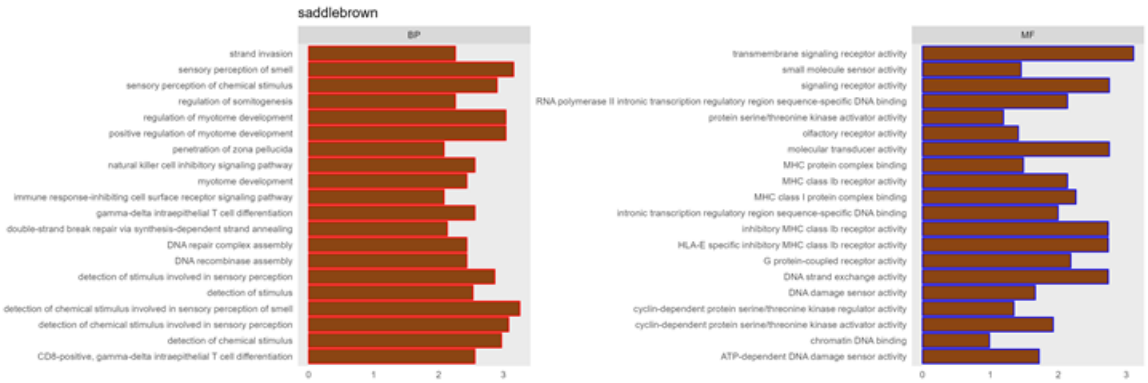
C



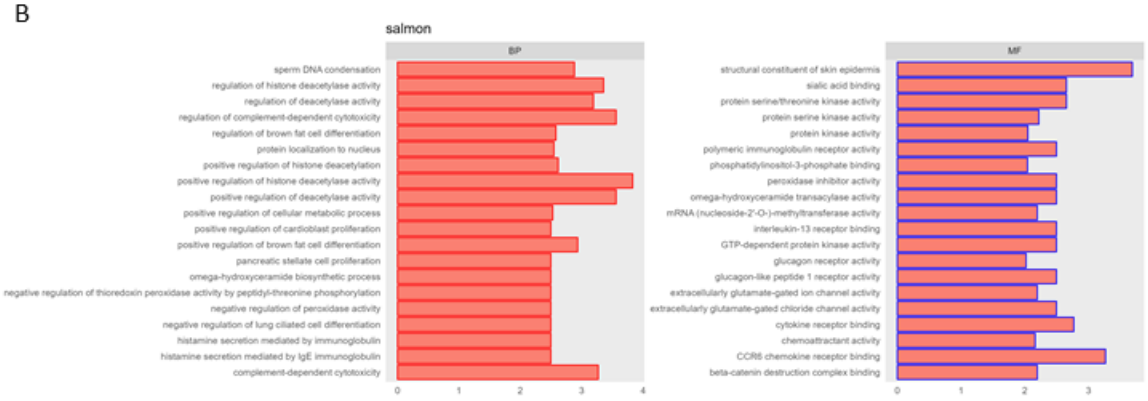
Supplementary Figure 17-15. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

Supplementary Figure 3-16

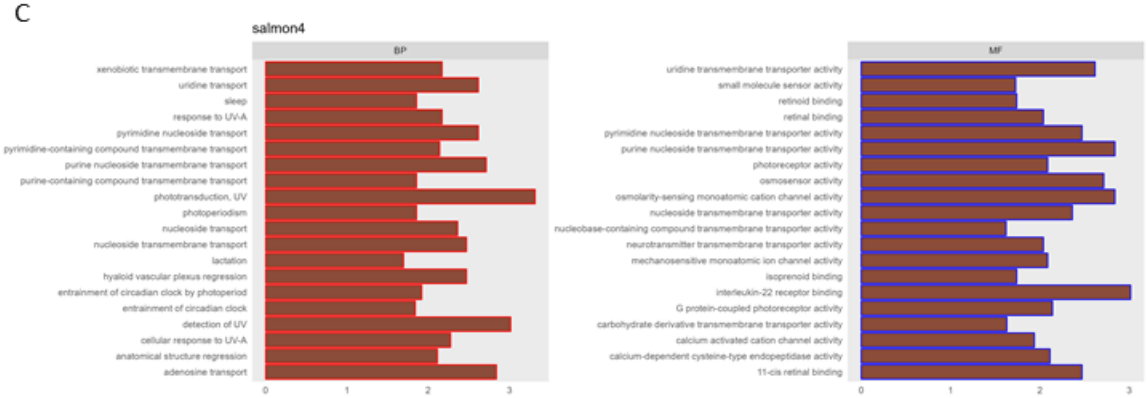
A



B

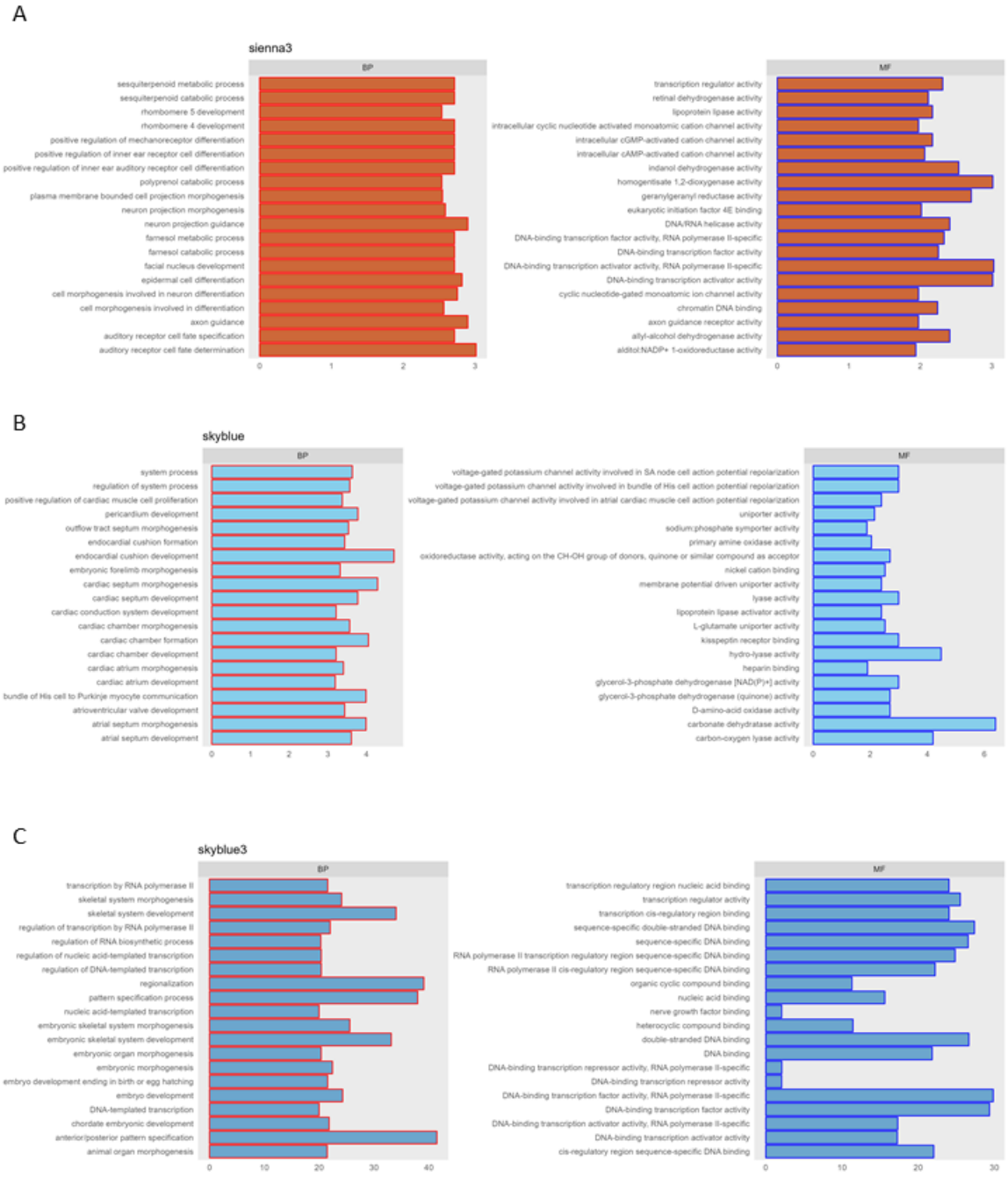


C



Supplementary Figure 18-16. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

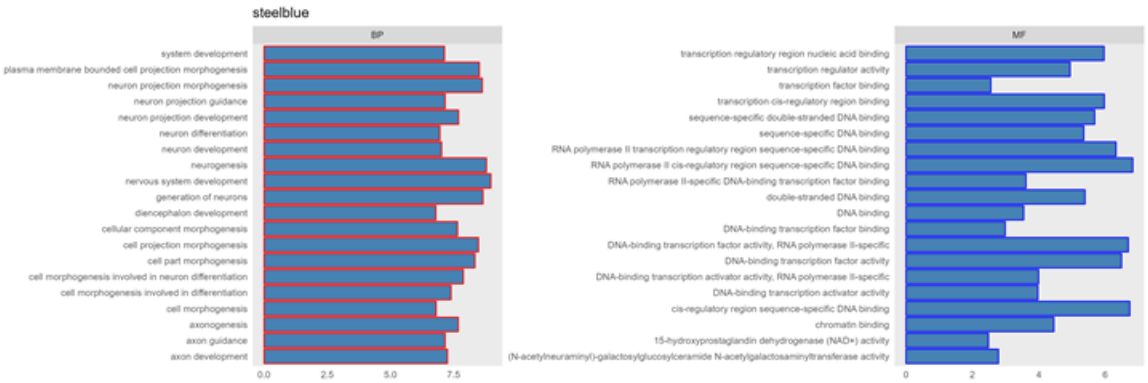
Supplementary Figure 3-17



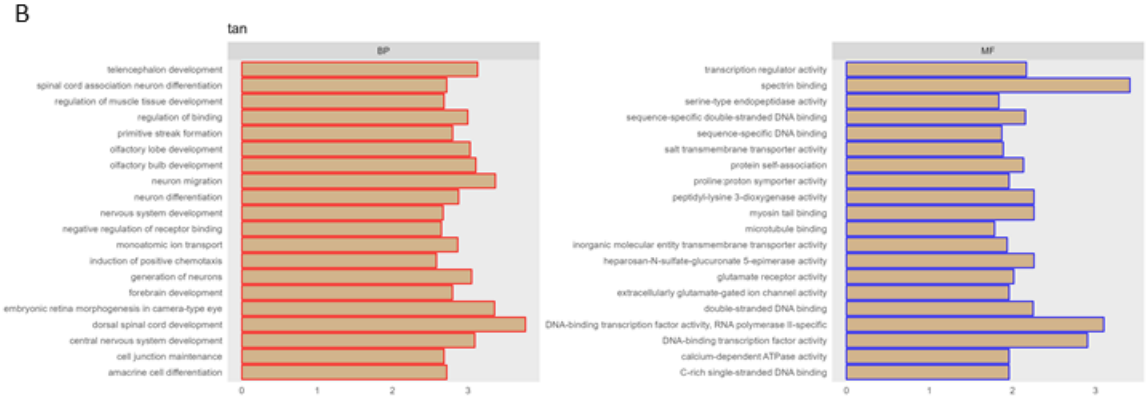
Supplementary Figure 19-17. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

Supplementary Figure 3-18

A



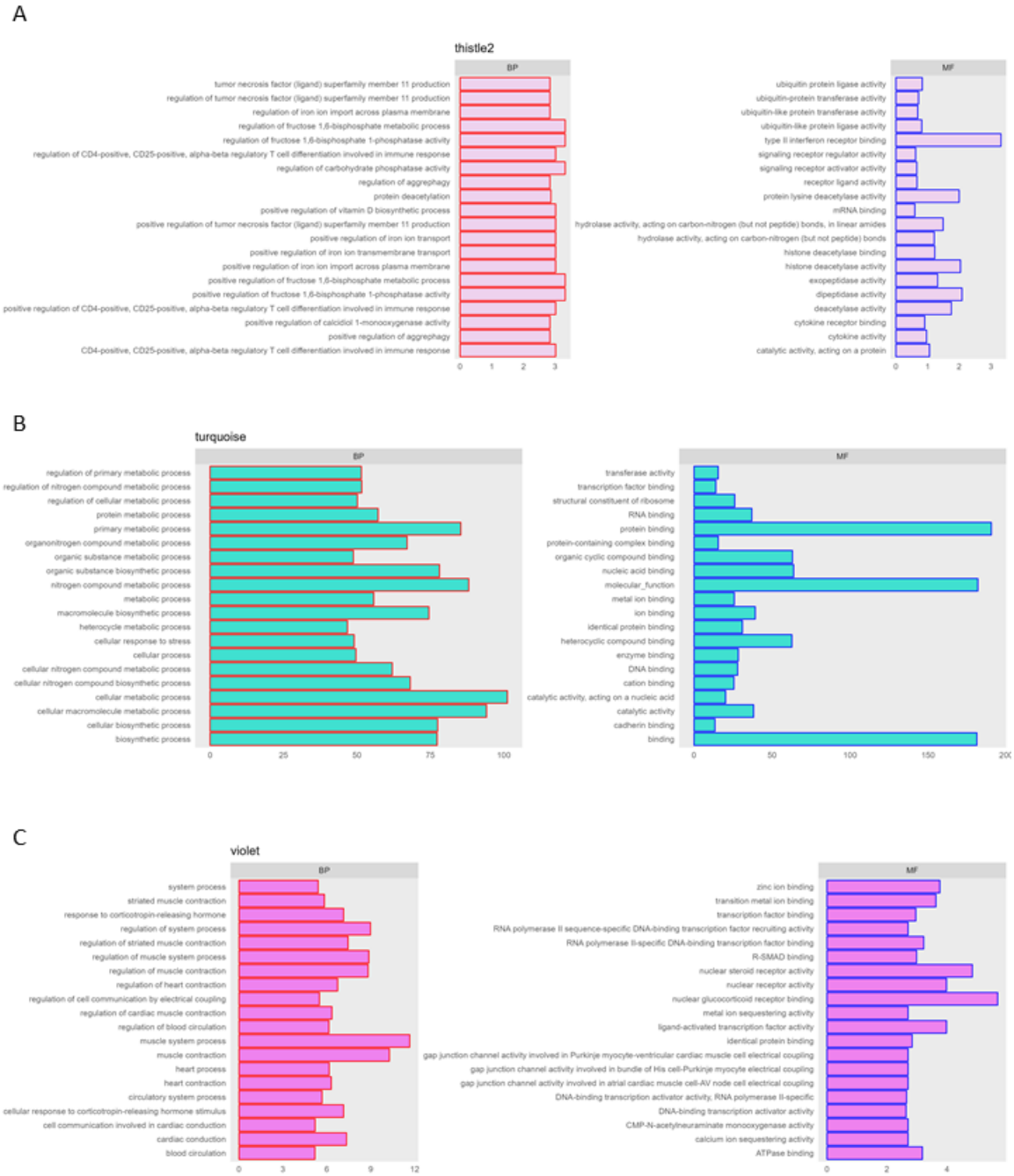
B



C

Supplementary Figure 20-18. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

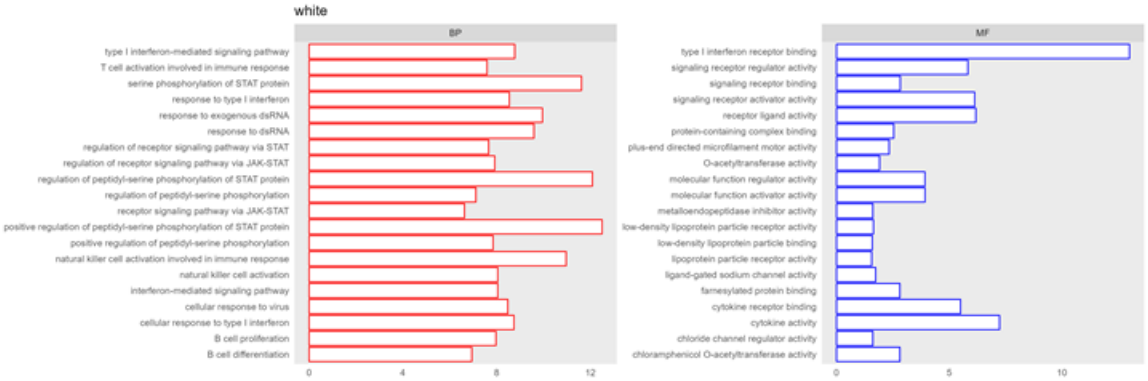
Supplementary Figure 3-19



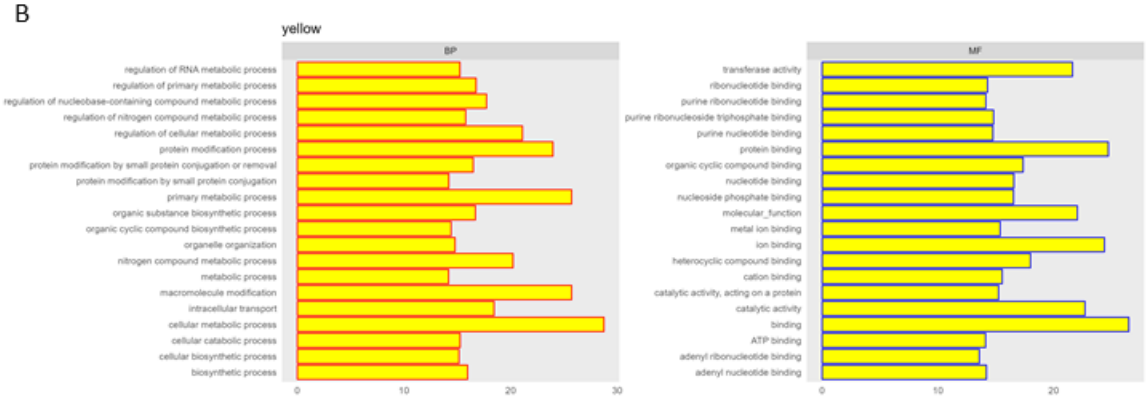
Supplementary Figure 21-19. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

Supplementary Figure 3-20

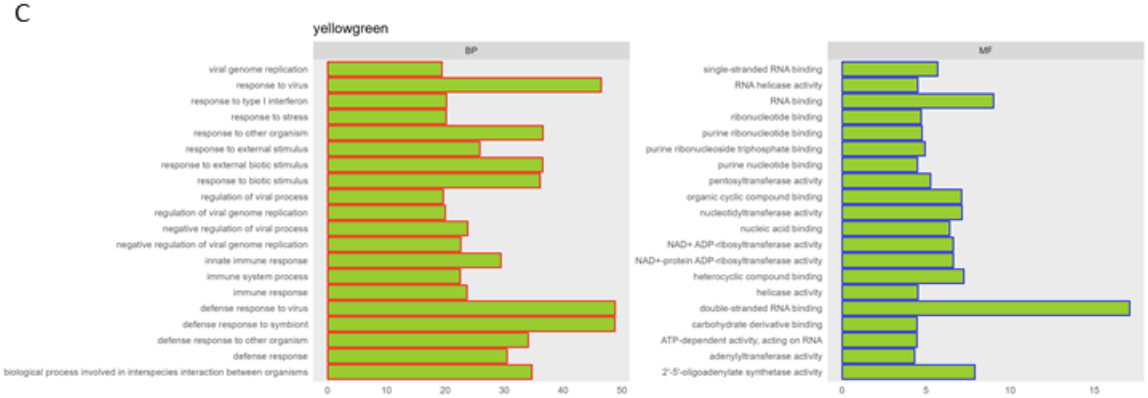
A



B



C



Supplementary Figure 22-20. WGCNA gene modules predict functions for genes where functions are not known. The statistical significance (-log10(p)) of biological processes (left) and molecular functions (right) are shown

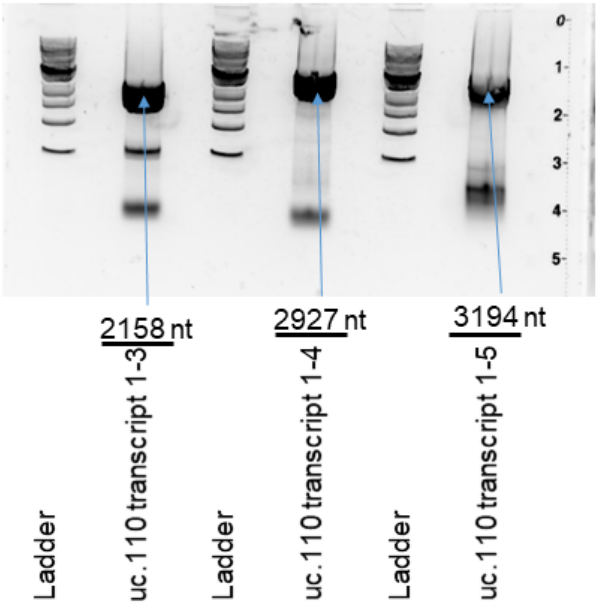
Supplementary Figure 4

uc.110

A

Chrom	Start	End	Length	ID
chr2	236,162,737	236,162,980	243	uc.110
chr2	236,161,295	236,164,636	3341	TUC110

B



C

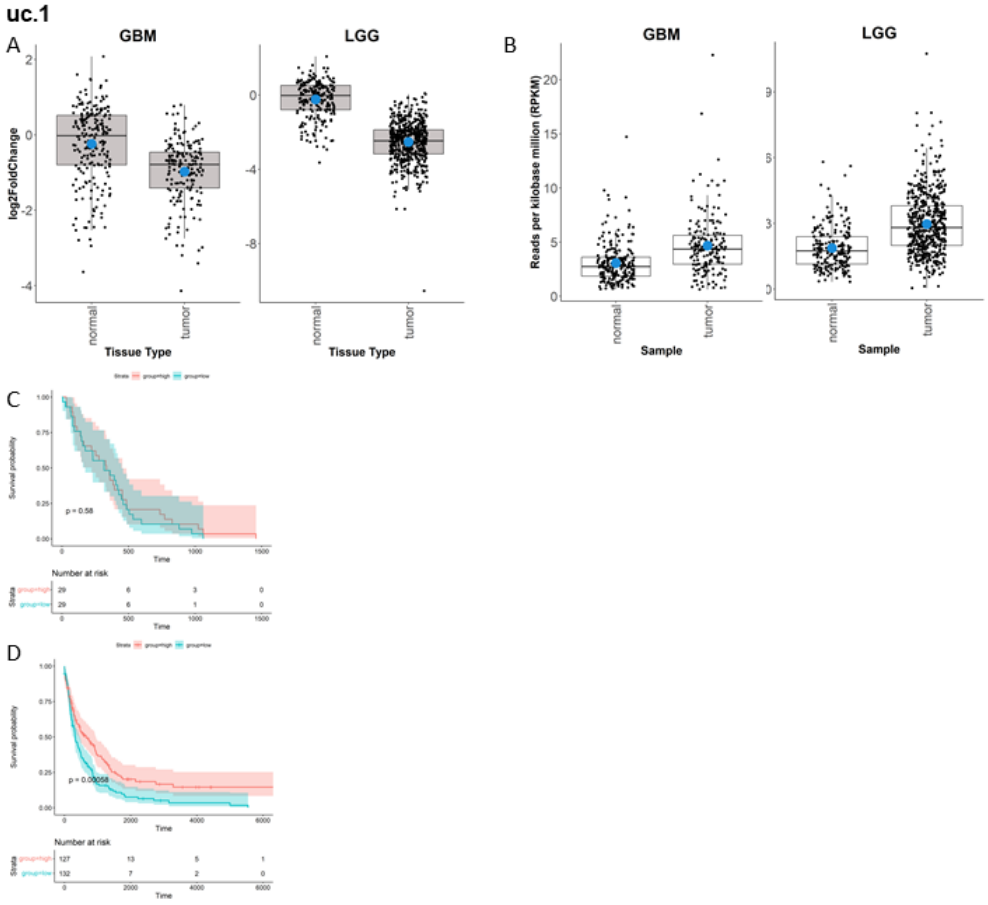
```
TCGGCTTAGGCCCCCACTTCTGAACTGTTCAAATAGACAATCGGTTTTGATGATGTAATTTATAATTTATGCAGGTT
GTCTACATACATATGGTGTTGCTTATGTGCGTTGGGGTTTCAGGTGCTTTATTACTCTGCCGGTGTGATTTGCATTC
ATGTTTGGTTACCAAGATCATATTTCTCCTGTTTACACCCTAAGGGCCTAAAATGAAGATGAATGCCACCTCCCTTC
GTCTGGGCTGTCCCTGTAGAACCCCGGCCTTAGCTTCCCCCAAGACCCCAGAGCCCTGCTACCCTGGAACTGGGCC
CTGAGCACTGCGCTGAAGCTGCCTGGGAGGACCCAGCGGCCAGAAGGTCGGAGCAGGCCACCTGCTCTGGCTGT
GCATGCCTGGGCAGATGCCTGGGGTCTCTGGGGACAGAAATAGGTGTGTGCACTGGAGCCTCCTTCCCTGGGAA
GCCACGGCTGGGCATAAGCGAGCCACTCTGCCACACCGCCACTGAAGTCCAGCAGGGGAAAGAAGAGGGCCCCG
AGAGGGTCTCTGCTGCCCGCGCTGCTGAATGCGTGGGCTCCAGCCACTTGGGACTAGGAAGGCCCCGAGGTGAG
AACCAGAGTGTGATCCAAAGGCTCCTGAGGCCGGAAAGCTAGAAGGTCACTCCCGGCTTTCAGTGCACCTAGAAC
CACCCGCCATGGGGCATAACATGGGACTCTGTTCCAGACAGGTCTGGGGCTGGAGGGAGCAGCCTTGGAAGCCC
TGTGCGAGCTTCACCAGATGCAGCTGAAATAATTAAGCCCATTAAGCTCAAACAAACAGAAACAAGAGGGAGAG
GGTGCTGACCCTGAGAGAAAGGGTGTCTTGGAATGCAGGTACCAGCGATGTGCCCAGGCTTGTACAAGCACCTA
GGCTGAAGGCCAGACGCACAGACAGCCATAGAGTCAGGATGTCCTCCGAGGCAAACCCGGGCTCAAAAAAGCCT
GGGGGGGCCACTCCCTGCTCTCCATTTAGCTGACAAGGCATCTGTGAAAGAACCACCCAGGGGGCTGAAGACAG
GAAGCAAAGGCTGGATGCACAAGCATGCACTAAGTCTCGAAAATTCCAGCGCAGGTTGGGATTTGAGCAGCTTCC
AGTGCAGGTGGGTCCCGGCTGCTGGGAAGGTAGTCCTGGGCTCCTTCTGGAAGGGGAAGTTGCTTGAGGCCCAG
GACCAAACTAAAGAGGGTCCTGCCCCAAGAACAGGGAGATGGCTGCGGATTTTCAGCATACCTTGGGTTCTCCAA
TGGGGCAGGGGTGCCTGGCTTGCCAGGGAGCAAAGGCAGTCACTGGGTGGAGTGACTGTTCTTGTGGCCCCCCT
CCGACCTCTGCCTGGGTCGGTGGCTGCGGGGTGCAGGCAGCCCTGAGGCGTCCCAGTGCACGGCTGGGGGGCGC
AGGTCCAAGGAGCTCCCAAGTGTAATTATTTGATctgcagatgtaagtggaatttattgtaacaacaaatatagtgatgtcaaaaccc
aaccttggattaaagccggcagccaaaggggaagtgtattaatttccaacagcatctcaggccagagcctattagaacaagctattcttcaggccccc
ccatcagaattaatcattgggagttaatttgaagctcagacaagttgctgttaatttagtgcagggaggacgggatggaataaaaagctgaggtgctg
gccgagcctcgtgggtctcattaatcagccagctgagacggcctggcttgattacaatttgcaaaaaaattcattagggcccgggcgattgacaattttt
ctctctgcctgtgatgtgactcattaggcagccaaatgaaagccatgatgacggccatgatgacagcggccatcaagcgctgcttgtttatcttcctccc
ctcggcctgtctcccgggccttggcctccctgccgggctctcgggggcgccccggggctcaggggttagggagggagggggtcagcgccccgaggag
caggagggagggcggcatcggccagcttggggagctcgggccctctcctcctcgtgcctccatagggcgaggacctcgaggctgggagggcccgga
gctcggtcccaggagcggggcttcctgacggcggcctccggttcctagtgggcttcgcttt
```

521

522 *Supplementary Figure 4. The uc.110 TUCR is part of a larger ncRNA transcript. A) We used de novo transcript*
523 *reassembly of TCGA glioma RNA-Seq data to identify the predicted sequence for the novel full RNA transcript*
524 *containing uc.110. Table depicts uc.110 ultraconserved and predicted full transcript genomic locations and length in*
525 *nucleotides (nt). B) PCR gel electrophoresis demonstrating validated uc.110 transcript variants. C) The validated full*
526 *sequence of the 2,158 nt uc.110 transcript is provided. The ultraconserved uc.110 region is colored red and primer*
527 *sequences are colored green.*

528

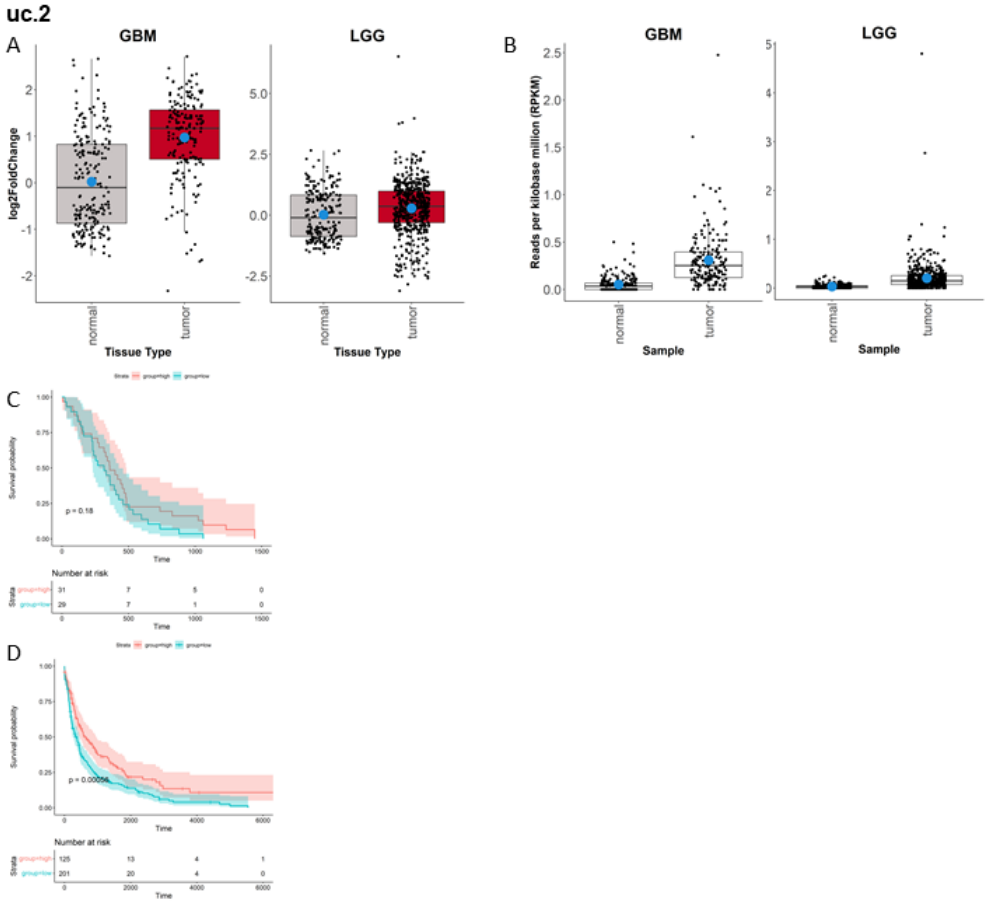
Supplementary Figures 5-1



529

530 *Supplementary Figures 5-1. An exploration of uc.1 in gliomas. A) Box- and dotplot showing TUCR deregulation in*
531 *GBM and LGG. Facets represent disease type. Red boxes represent upregulated TUCRs. Green boxes represent*
532 *downregulated TUCRs. Gray boxes represent TUCRs that are not deregulated. B) Box- and dotplot showing TUCR*
533 *absolute expression in GBM and LGG. Facets represent disease type. C) Kaplan-Meier showing TUCR association*
534 *with GBM prognosis. Red line represents the TUCR high expression group. Teal line represents the TUCR low*
535 *expression group. D) Kaplan-Meier showing TUCR association with LGG prognosis. Red line represents the TUCR*
536 *high expression group. Teal line represents the TUCR low expression group.*

Supplementary Figures 5-2

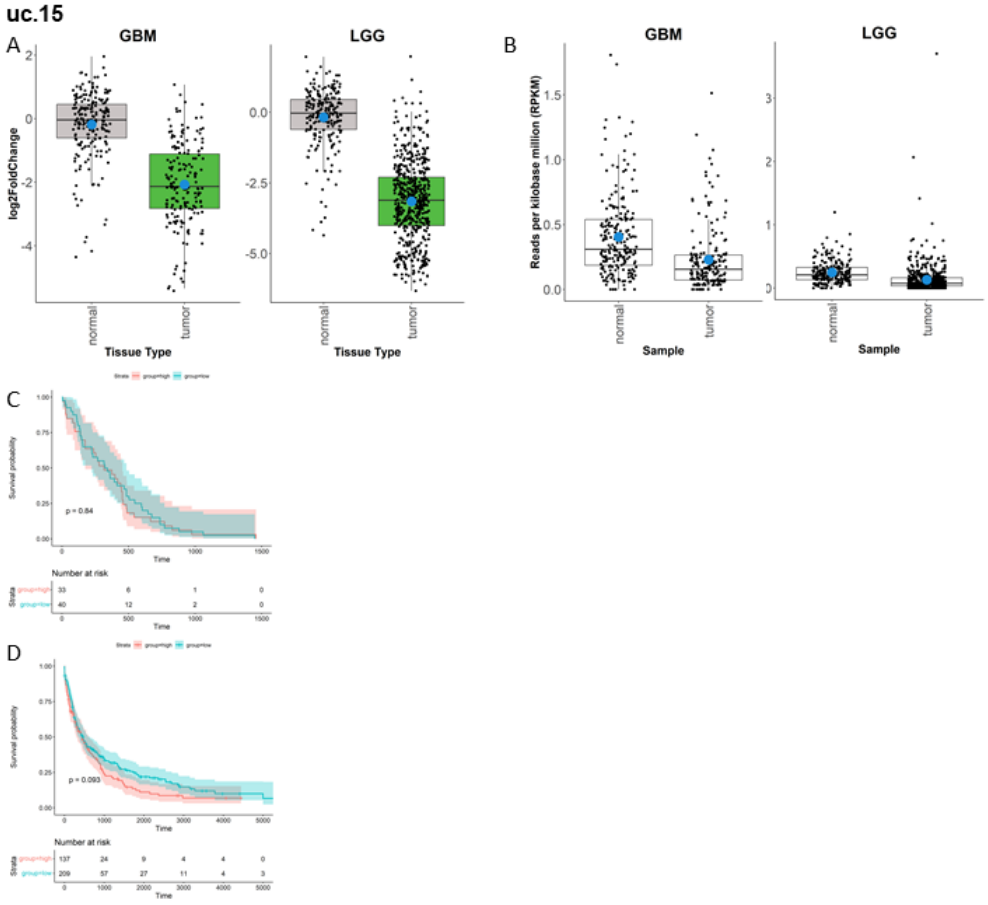


537

538 *Supplementary Figures 5-2. An exploration of uc.2 in gliomas. A) Box- and dotplot showing TUCR deregulation in*
539 *GBM and LGG. Facets represent disease type. Red boxes represent upregulated TUCRs. Green boxes represent*
540 *downregulated TUCRs. Gray boxes represent TUCRs that are not deregulated. B) Box- and dotplot showing TUCR*
541 *absolute expression in GBM and LGG. Facets represent disease type. C) Kaplan-Meier showing TUCR association*
542 *with GBM prognosis. Red line represents the TUCR high expression group. Teal line represents the TUCR low*
543 *expression group. D) Kaplan-Meier showing TUCR association with LGG prognosis. Red line represents the TUCR*
544 *high expression group. Teal line represents the TUCR low expression group.*

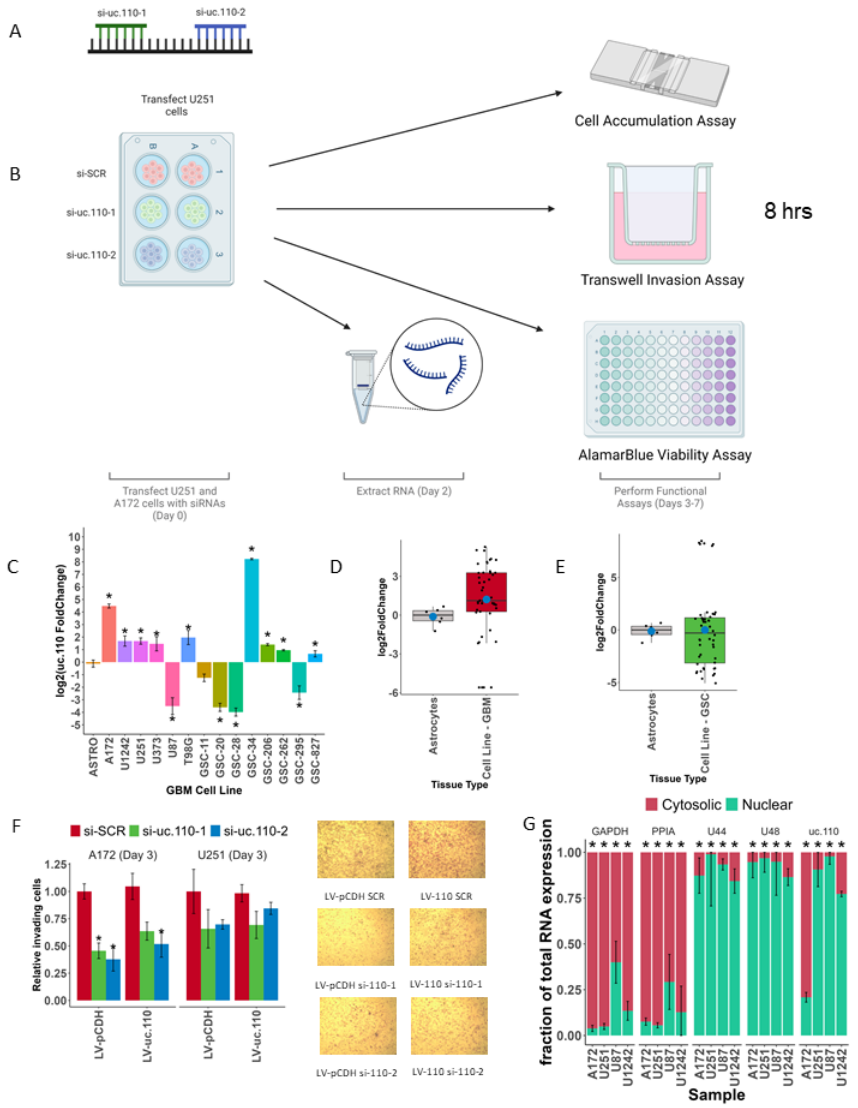
545

Supplementary Figures 5-15

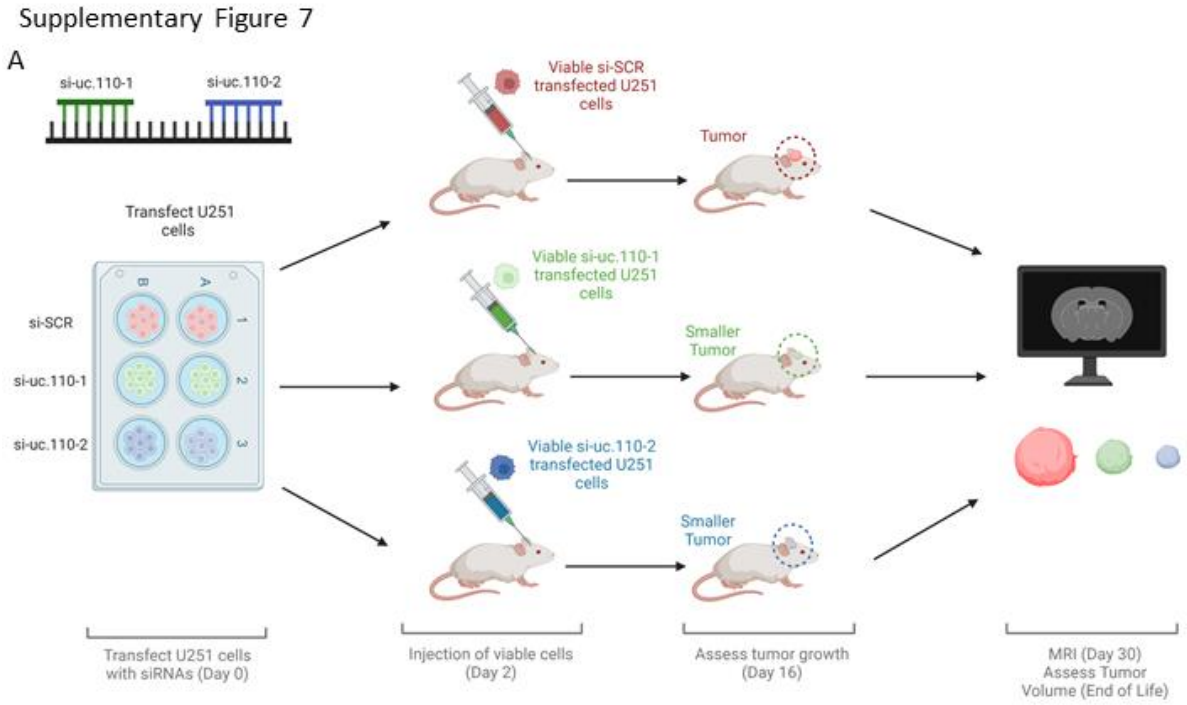


Supplementary Figures 5-15. An exploration of uc.15 in gliomas. A) Box- and dotplot showing TUCR deregulation in GBM and LGG. Facets represent disease type. Red boxes represent upregulated TUCRs. Green boxes represent downregulated TUCRs. Gray boxes represent TUCRs that are not deregulated. B) Box- and dotplot showing TUCR absolute expression in GBM and LGG. Facets represent disease type. C) Kaplan-Meier showing TUCR association with GBM prognosis. Red line represents the TUCR high expression group. Teal line represents the TUCR low expression group. D) Kaplan-Meier showing TUCR association with LGG prognosis. Red line represents the TUCR high expression group. Teal line represents the TUCR low expression group.

Supplementary Figure 6

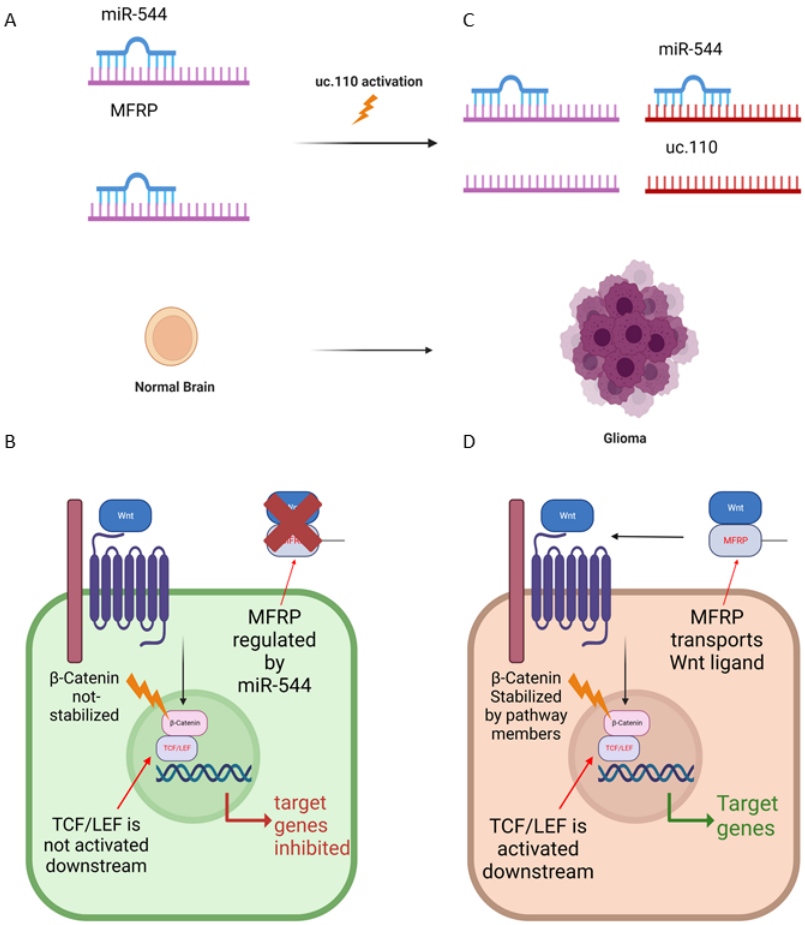


*Supplementary Figure 6. The uc.110 TUCR operates as an oncogene (cont.) A) Cartoon depicting two siRNAs that target different regions of the uc.110 TUCR. One starts at nt 96/243 (blue), and the other at nt 195/243 (green). B) Cartoon schematic depicting transfection protocol. Cells were transfected with siRNAs using Lipofectamine 2000 at D0. RNA was collected at D2, and functional assays were performed from D3-D7. C) Bar graph depicting uc.110 upregulation in banked GBM cell lines. D) Boxplot representing uc.110 expression in pooled glioma adherent cell lines versus normal human astrocytes. Red boxes indicate an upregulated TUCR. D) Boxplot representing uc.110 expression in pooled glioma adherent cell lines versus normal human astrocytes. Red boxes indicate an upregulated TUCR. E) Boxplot representing uc.110 expression in pooled glioma stem cell lines versus normal human astrocytes. Green boxes indicate a downregulated TUCR. F) Bar graph depicting that the cell invasion phenotype is rescued in A172 and U251 cells with uc.110 overexpression in the presence of siRNAs. Images are representative of the listed sample. si-SCR = scrambled control siRNA (red), si-uc.110-1 = siRNA targeting uc.110 at nucleotide 96/243 (green), si-uc.110-2 = siRNA targeting uc.110 at nucleotide 195/243 (blue). G) Cell fractionation bar graph depicting that uc.110 is a predominantly nuclear RNA molecule, with some cytosolic expression in A172s cells. Facets represent cytosolic (red) control genes (GAPDH,PPIA), nuclear (teal) control genes (U44,U48), and the uc.110 TUCR. * = p < 0.05*



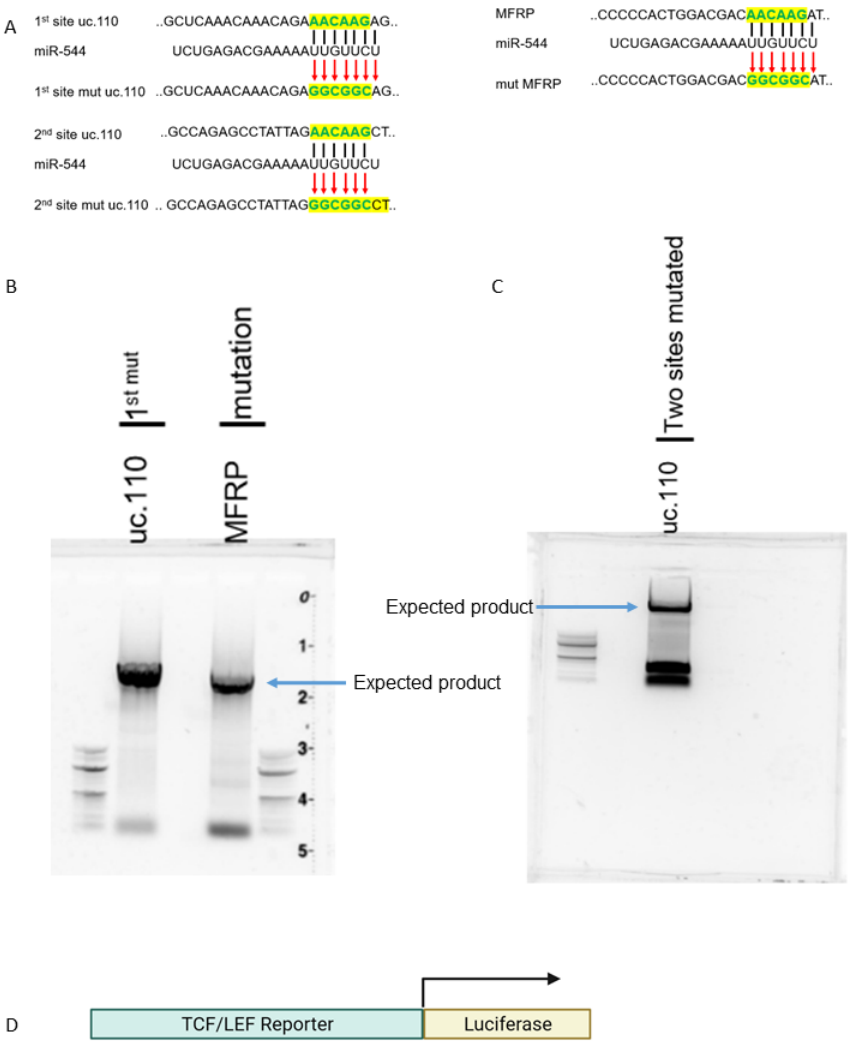
Supplementary Figure 7. The uc.110 TUCR promotes tumor growth in vivo A) Cartoon depiction of mouse experiment workflow. Cells with transfected with siRNAs using Lipofectamine 2000 at D0. RNA was extracted and cells with injected into mice at D2. Tumor growth was assessed weekly, starting at D16, via MRI.

Supplementary Figure 8



Supplementary Figure 8. The uc.110 TUCR activates Wnt-signaling by sponging miR-544 from membrane frizzled related protein (MFRP) 3'UTR. A) Schematic depicting model for miR-544 sponging by uc.110. B) Schematic depicting simplified repressed Wnt-signaling pathway. In the normal brain, MFRP is regulated by miR-544 as depicted in 6A. C) Activation of uc.110 in glioma tumors leads to decreased bioavailability of miR-544. This increases the bioavailability of MFRP. D) Schematic depicting simplified activated Wnt-signaling pathway. When MFRP bioavailability is increased by uc.110 activation, as depicted in 6C, Wnt-signaling is also increased.

Supplementary Figure 9



Supplementary Figure 9. Luciferase assays confirm that uc.110 binds miR-544 and regulates wnt-signaling through MFRP. A) Schematic depiction of TCF/LEF luciferase reporter construct used to measure downstream Wnt-signaling pathway activation. TCF/LEF binds to the reporter region (green) and activates luciferase (yellow). B) Schematic depicting miR-544 binding site mutations for uc.110 (two sites) and MFRP (one site). Top row represents wild-type binding sites. Middle row is the miR-544 binding region. Bottom row are mutated sites. C) PCR gel showing expected products from uc.110 (first site) and MFRP mutations. D) PCR gel showing expected product from the second miR-544 binding site in uc.110.

MATERIALS AND METHODS

Data Availability Statement

RNA-Seq data for Figure 6A will be made available on the Gene Expression Omnibus (GEO) prior to publication. Detailed methods, including access information for all datasets used, can be found in a repository at: **github.com/abounaderlab/tucr_project.** Detailed TUCR results can be found at **abounaderlab.org/tucr-database**. Please refer to the corresponding author for any data access questions.

TUCR Annotations [29, 30]

TUCR annotations were performed manually by overlaying consensus TUCR genomic annotation tracks to the hg38 human genome in the UCSC Genome Browser. In parallel, bedtools closest was used to identify genes that are intergenic or intragenic. These results were then cross referenced to identify a consensus genomic annotation for each TUCR.

TUCR Chromatin Landscaping

U87 H3K4me3, RNA Pol.II, and H3K27ac CHIP-Seq data and U87 ATAC-Seq data were acquired from the Gene Expression Omnibus. Randomized control TUCRs were generated using Quinlan Labs' bedtools [31, 32] and the shuffle command.[31, 32] Bedtools fisher and R/RStudio [53, 54] were used to perform chi-square tests to compare predicted overlaps of peaks to expected peaks.

TCGA AND GTex RNA-Seq Data [33, 34]

GBM (n = 161) and LGG (n = 505) RNA-Seq data were acquired from the Cancer Genome Atlas and were compared to normal brain cortex from the Genotype-Tissue Expression Database (GTex, n = 260) using a workflow including bedtools, bowtie, the SRA toolkit, and R/RStudio.

TUCR Expression, Deregulation, and Survival Analyses [33, 34, 53, 54]

TUCR expression, deregulation, and survival analyses, were analyzed using processed TCGA and GTex RNA-Seq data and a workflow using R/RStudio.

TUCR weighted gene correlation network analysis (WGCNA) [36]

TUCR WGCNA was performed using processed TCGA and GTex RNA-Seq data using a modified version of the R/RStudio workflow designed by Drs. Peter Langfelder and Steve Horvath at UC Los Angeles.

De novo transcript reassembly and validation [35]

De novo transcript assembly was performed on TCGA GBM and LGG RNA-Seq data using standard protocols and the *stringtie* bioinformatics package. Results were validated using PCR: 10 min at 95°C, followed by 40 cycles of 10 seconds at 95°C and 1 minute at 60°C.

Patient Samples

GBM Tumor samples were acquired from the UVA Tumor Bank. Detailed patient information can be found as a supplement (UVATumorBank_data.csv).

Cell Lines

U87, U251, A172, and T98G glioblastoma cell lines were used in *in vitro* experiments and were acquired from ATCC. U87 cells were cultured in 500 mL minimum essential media (MEM) Earles (Gibco, #.11095-080) containing 5 mL penicillin/streptomycin (pen/strep, Gibco, Cat #.15140-133), 5 mL MEM non-essential amino acids (NEAA, Gibco, #.11140-050), 5 mL sodium pyruvate (Gibco, 100 nM, #.11360-070), 10 mL sodium bicarbonate (Gibco, 7.5%, #.25080-094), and 50 mL fetal bovine serum (FBS). T98G cells were cultured in 500 mL MEM Earles media containing 5 mL pen/strep, 5 mL NEAA, 5 mL sodium pyruvate, and 50 mL FBS. A172 cells were cultured in 500 mL Dulbecco's modified eagle media (DMEM, Gibco, #.11965-092) containing 5 mL pen/strep, and 50 mL FBS. U251 cells were cultured in 500 mL RPMI L-Glutamine media (Gibco, #.11875-093) containing 5 mL pen/strep and 25 mL FBS. GSC-34 and GSC-28 glioblastoma stem cells cells were cultured in neurobasal (L-glutamine negative) media (Gibco, #.21103-049) containing 5 mL pen/strep, 5 mL B-27 (without Vit-A, Gibco, #.12587-010), 2.5 mL N-2 (Gibco, #.17502-048), 1 mL EGF, 1 mL FGF, and 1.25 mL L-Glutamine. All cell media contained in 5 µL Plasmocure reagent to prevent mycoplasma contamination.

Primer and Oligo Design

Primers and siRNAs were designed using the Primer3 and Thermofisher design portals, respectively. uc.110 forward primer sequence is 5'-CAGCCAAAGGGGAAGTGTAT-3', and the reverse sequence is 5'-CCGTCCTCCCTGCACTAAAT-3'.

MFRP forward primer sequence is 5'- GCATCTATTCATGTGGCAGGC-3', and the reverse sequence is 5'- TACTCCGGACCCTCCAGTTG-3'.

The miR-544 precursor was ordered from Invitrogen (#.AM17100). Negative control oligos were ordered from Ambion (#.AM4635).

uc.110 stable overexpression

The full uc.110 transcript from "de novo transcript reassembly and validation" was cloned into the pCDH-EF1-MCS-BGH-PGK-GFP-T2A-Puro vector (# CD550A-1) using stbl3 competent e.coli cells and ampicillin selection. Amplified vector was extracted using the miniprep kit (Qiagen, #.27106). 0.75 µg of this vector, 0.75 µg of psPAX2 lentiviral gag-pol packaging vector, and 0.5 µg of pMD.2G VSV-G enveloping protein was transfected in 6 µL X-tremeGENE transfection reagent (#.06366236001) into 293T cells per manufacturer instructions to generate a lentivirus that was transduced to U87, U251, and A172 cells in media without antibiotics. These cells were subjected to antibody (puromycin) selection for uc.110-positive cells at D3.

uc.110 quantitative (q)PCR

Total RNA was isolated using the RNEasy+ kit (Qiagen, #.74134) according to manufacturer instructions. RNA concentration and purity was measured via nanodrop. 800 ng of cDNA was synthesized (BIORAD T100 Thermal Cycler) using the iScript (BIORAD, #. 1708890) synthesis kit per manufacturer instructions. A 20 µL reaction mixture was then created for each condition with the following concentrations: 1 µL of combined forward/reverse primers (5 µM), 10 µL of iQ SYBR Green master mix (#1798880), 4 µL of nuclease free water, and 5 µL of synthesized cDNA. These reactions were cycled (BIORAD CFX Real Time System) in 96-well plates: 10 min at 95°C, followed by 40 cycles of 10 seconds at 95°C and 1 minute at 60°C.

Cell Counting (Accumulation) Assay [37-39, 44]

Cells were seeded in 6-well culture plates with full serum media at 30,000/well density at D-1. At D0, each well was transfected via master mix 3 µL of siRNAs (20 µM) via 9 µL Lipofectamine 2000 (Invitrogen, #.11668-019) in 300 µL OPTI-MEM (Gibco, #.31985-070) and 700 µL antibiotic and empty media for 6 hours. At 6 hours, media were replaced with fresh media containing antibiotics and FBS. Cells were then counted via haemocytometer at Days 1, 3, 5, and 7 for each cell line.

Transwell Cell Invasion/Migration Assay [42-44]

Cells were seeded in 6-well culture plates with full serum media at 300k/well density at D-1. At D0, each well was transfected via master mix 3 µL of siRNAs (20 µM) via 9 µL Lipofectamine 2000 in 300 µL OPTI-MEM and 700 µL antibiotic and empty media for 6 hours. At 6 hours, the media were replaced with fresh media containing antibiotics and FBS. The cells were then seeded in empty media at 200k/chamber into Transwell Invasion Chambers coated with Collagen IV. After 8 hours, non-invading cells were cleared and invading cells were stained with Crystal Violet.

AlamarBlue Cell Viability Assay {Voytik-Harbin, 1998 #21645;Kumar, 2018 #21629}

Cells were seeded in 96-well culture plates with full serum media at 10k/well density at D-1. Border wells were filled with media to account for edge effects. At D0, each well was transfected via master mix 1 µL of siRNAs (20 µM) via 3 µL Lipofectamine 2000 in 30 µL OPTI-MEM and 70 µL antibiotic and empty media for 6 hours. At 6 hours, media were replaced with fresh media containing antibiotics and FBS. Functional assays were performed using the AlamarBlue kit (Life Technologies #.A50100) per manufacturer instructions. Reactions were allowed to proceed for 1 hour.

Ex vivo knockdown of uc.110

Cells were seeded in 6-well culture plates with full serum media at 300k/well density at D-1. At D0, each well was transfected with 3 µL of siRNAs (20 µM) via 9 µL Lipofectamine 2000 in 300 µL OPTI-MEM and 700 µL antibiotic and empty media for 6 hours. At 6 hours, the media were replaced with fresh media containing antibiotics and FBS. Mouse experiments were performed

using xenograft models and intracranial injections of U251 cells post transfection with siRNA oligonucleotides. Cells were injected at D2 and were imaged at two-week intervals via MRI. Survival was assessed sured daily and tumor volume was measured at the end of life.

Characterization of transcriptome post-uc.110 knockdown

Cells were seeded in 6-well culture plates with full serum media at 300k/well density at D-1. At D0, each well was transfected with 3 µL of siRNAs (20 µM) using 9 µL Lipofectamine 2000 in 300 µL OPTI-MEM and 700 µL antibiotic and empty media for 6 hours. At 6 hours, the media were replaced with fresh media containing antibiotics and FBS. RNA Libraries were collected and sequenced via RNA-Seq at D2.

Luciferase Assays

Cells were seeded in 6-well culture plates with full serum media at 300k/well density at D-1. At D0, each well was transfected with 3 µL of siRNA/miRNA (20 µM) using 9 µL Lipofectamine 2000 in 300 µL OPTI-MEM and 700 µL antibiotic and empty media for 6 hours. At 6 hours, the media were replaced with fresh media containing antibiotics and FBS. MFRP and uc.110 sequences were cloned into the PROMEGA pmirGLO Luciferase vector (E1330). BPS Dual reporter luciferase assays were ordered for TCF/LEF (#.60500) and uc.110/MFRP (#.60683) experiments.

Statistical Analyses

Comparisons between means of samples were performed using Student's t-test and one-way ANOVAs. Comparisons between categorical variables were performed using chi-squared and Fisher's exact test. Comparisons were considered statistically significant if the p-value was less than 0.05. Molecular experiment tests were performed in SigmaPlot 14.0, while computational experiment tests were performed using bedtools and/or RStudio. Detailed methods for all analyses are provided as outlined in the data availability statement.

AUTHOR CONTRIBUTIONS

Author contributions are defined using Elsevier's CRediT format:

Myron Gibert Jr: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing, Visualization, Supervision, Project Administration, Funding Acquisition

Ying Zhang: Methodology, Investigation, Resources, Formal Analysis, Data Curation, Review and Editing

Shekhar Saha: Methodology, Investigation, Resources, Formal Analysis, Data Curation, Review and Editing

Pawel Marcinkiewicz: Investigation, Resources, Formal Analysis, Data Curation, Review and Editing

Sylwia Bednarek: Methodology, Validation, Formal Analysis, Investigation, Review and Editing, Visualization

Collin Dube: Methodology, Writing, Visualization, Review and Editing

Kadie Hudson: Methodology, Writing, Visualization, Review and Editing

Yunan Sun: Methodology, Writing, Visualization, Review and Editing

Bilhan Chagari: Formal Analysis, Investigation, Writing, Data Curation

Aditya Sarkar: Formal Analysis, Investigation, Writing, Data Curation

Christian Roig-Laboy: Formal Analysis, Investigation, Writing, Data Curation

Natalie Neace: Formal Analysis, Investigation, Writing, Data Curation

Karim Saoud: Formal Analysis, Investigation, Writing, Data Curation

Initha Setiady: Formal Analysis, Investigation, Writing, Data Curation

Farina Hanif: Methodology, Investigation, Review and Editing

Pankaj Kumar: Methodology, Software, Validation, Formal Analysis, Resources, Data Curation

Benjamin Kefas: Methodology, Investigation, Review and Editing

Roger Abounader: Conceptualization, Methodology, Resources, Review and Editing, Supervision, Project Administration, Funding Acquisition

ACKNOWLEDGMENTS

This article was supported by NIH UO1 CA220841, NINDS 1R21NS122136-01, NINDS RO1 NS122222, NCI Cancer Center Support Grant 5P30CA044579, and a Schiff Foundation grant (all to R.A.). Special thanks goes to Drs. Stephen Turner and Alex Koeppel for their work with the Bioinformatics Core and early contributions to many the computational experiments in this study.

REFERENCES

1. Bejerano, G., et al., *Ultraconserved elements in the human genome.* Science, 2004. **304**(5675): p. 1321-5.
2. Gibert, M.K., Jr., et al., *Transcribed Ultraconserved Regions in Cancer.* Cells, 2022. **11**(10).
3. Calin, G.A. and C.M. Croce, *Chronic lymphocytic leukemia: interplay between noncoding RNAs and protein-coding genes.* Blood, 2009. **114**(23): p. 4761-70.
4. Calin, G.A., et al., *Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas.* Cancer Cell, 2007. **12**(3): p. 215-29.
5. Colwell, M., et al., *Evolutionary conservation of DNA methylation in CpG sites within ultraconserved noncoding elements.* Epigenetics, 2018. **13**(1): p. 49-60.
6. Edwards, J.K., et al., *MicroRNAs and ultraconserved genes as diagnostic markers and therapeutic targets in cancer and cardiovascular diseases.* J Cardiovasc Transl Res, 2010. **3**(3): p. 271-9.
7. Fabris, L. and G.A. Calin, *Understanding the Genomic Ultraconservations: T-UCRs and Cancer.* Int Rev Cell Mol Biol, 2017. **333**: p. 159-172.
8. Lujambio, A., et al., *CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer.* Oncogene, 2010. **29**(48): p. 6390-401.
9. McCole, R.B., et al., *Abnormal dosage of ultraconserved elements is highly disfavored in healthy cells but not cancer cells.* PLoS Genet, 2014. **10**(10): p. e1004646.
10. Scaruffi, P., *The transcribed-ultraconserved regions: a novel class of long noncoding RNAs involved in cancer susceptibility.* ScientificWorldJournal, 2011. **11**: p. 340-52.
11. Zambalde, E.P., et al., *Transcribed Ultraconserved Regions Are Associated with Clinicopathological Features in Breast Cancer.* Biomolecules, 2022. **12**(2).
12. Braicu, C., et al., *The Function of Non-Coding RNAs in Lung Cancer Tumorigenesis.* Cancers (Basel), 2019. **11**(5).

13. Catana, C.S., et al., *Non-coding RNAs, the Trojan horse in two-way communication between tumor and stroma in colorectal and hepatocellular carcinoma.* Oncotarget, 2017. **8**(17): p. 29519-29534.
14. Irimie, A.I., et al., *The Unforeseen Non-Coding RNAs in Head and Neck Cancer.* Genes (Basel), 2018. **9**(3).
15. Ling, H., et al., *Junk DNA and the long non-coding RNA twist in cancer genetics.* Oncogene, 2015. **34**(39): p. 5003-11.
16. Liz, J., et al., *Regulation of pri-miRNA processing by a long noncoding RNA transcribed from an ultraconserved region.* Mol Cell, 2014. **55**(1): p. 138-47.
17. Sakamoto, N., et al., *Non-coding RNAs are promising targets for stem cell-based cancer therapy.* Noncoding RNA Res, 2017. **2**(2): p. 83-87.
18. Kiran, M., et al., *A Prognostic Signature for Lower Grade Gliomas Based on Expression of Long Non-Coding RNAs.* Mol Neurobiol, 2019. **56**(7): p. 4786-4798.
19. Dey, B.K., A.C. Mueller, and A. Dutta, *Long non-coding RNAs as emerging regulators of differentiation, development, and disease.* Transcription, 2014. **5**(4): p. e944014.
20. Cruickshanks, N., et al., *Role and Therapeutic Targeting of the HGF/MET Pathway in Glioblastoma.* Cancers (Basel), 2017. **9**(7).
21. Davis, M.E., *Glioblastoma: Overview of Disease and Treatment.* Clin J Oncol Nurs, 2016. **20**(5 Suppl): p. S2-8.
22. Lathia, J.D., et al., *Cancer stem cells in glioblastoma.* Genes Dev, 2015. **29**(12): p. 1203-17.
23. Kalkan, R., *Glioblastoma Stem Cells as a New Therapeutic Target for Glioblastoma.* Clin Med Insights Oncol, 2015. **9**: p. 95-103.
24. Thakkar, J.P., et al., *Epidemiologic and molecular prognostic review of glioblastoma.* Cancer Epidemiol Biomarkers Prev, 2014. **23**(10): p. 1985-96.
25. Mao, H., et al., *Deregulated signaling pathways in glioblastoma multiforme: molecular mechanisms and therapeutic targets.* Cancer Invest, 2012. **30**(1): p. 48-56.
26. Altaner, C., *Glioblastoma and stem cells.* Neoplasma, 2008. **55**(5): p. 369-74.
27. Mourad, P.D., et al., *Why are systemic glioblastoma metastases rare? Systemic and cerebral growth of mouse glioblastoma.* Surg Neurol, 2005. **63**(6): p. 511-9; discussion 519.
28. Zhang, Y., et al., *The p53 Pathway in Glioblastoma.* Cancers (Basel), 2018. **10**(9).

29. Raney, B.J., et al., *Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser.* Bioinformatics, 2014. **30**(7): p. 1003-5.
30. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.
31. Quinlan, A.R., *BEDTools: The Swiss-Army Tool for Genome Feature Analysis.* Curr Protoc Bioinformatics, 2014. **47**: p. 11 12 1-34.
32. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.
33. Cancer Genome Atlas Research, N., et al., *The Cancer Genome Atlas Pan-Cancer analysis project.* Nat Genet, 2013. **45**(10): p. 1113-20.
34. Stanfill, A.G. and X. Cao, *Enhancing Research Through the Use of the Genotype-Tissue Expression (GTEx) Database.* Biol Res Nurs, 2021. **23**(3): p. 533-540.
35. Pertea, M., et al., *CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise.* Genome Biol, 2018. **19**(1): p. 208.
36. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC Bioinformatics, 2008. **9**: p. 559.
37. Hanif, F., et al., *miR-3174 Is a New Tumor Suppressor MicroRNA That Inhibits Several Tumor-Promoting Genes in Glioblastoma.* Int J Mol Sci, 2023. **24**(11).
38. Mulcahy, E.Q.X., et al., *MicroRNA 3928 Suppresses Glioblastoma through Downregulation of Several Oncogenes and Upregulation of p53.* Int J Mol Sci, 2022. **23**(7).
39. Cruickshanks, N., et al., *Discovery and Therapeutic Exploitation of Mechanisms of Resistance to MET Inhibitors in Glioblastoma.* Clin Cancer Res, 2019. **25**(2): p. 663-673.
40. Kumar, P., A. Nagarajan, and P.D. Uchil, *Analysis of Cell Viability by the alamarBlue Assay.* Cold Spring Harb Protoc, 2018. **2018**(6).
41. Voytik-Harbin, S.L., et al., *Application and evaluation of the alamarBlue assay for cell growth and survival of fibroblasts.* In Vitro Cell Dev Biol Anim, 1998. **34**(3): p. 239-46.
42. Stoellinger, H.M. and A.R. Alexanian, *Modifications to the Transwell Migration/Invasion Assay Method That Eases Assay Performance and Improves the Accuracy.* Assay Drug Dev Technol, 2022. **20**(2): p. 75-82.
43. Liu, X. and X. Wu, *Utilizing Matrigel Transwell Invasion Assay to Detect and Enumerate Circulating Tumor Cells.* Methods Mol Biol, 2017. **1634**: p. 277-282.
44. Justus, C.R., et al., *In vitro cell migration and invasion assays.* J Vis Exp, 2014(88).

45. Guessous, F., et al., *Cooperation between c-Met and Focal Adhesion Kinase Family Members in Medulloblastoma and Implications for Therapy.* Mol Cancer Ther, 2012. **11**(2): p. 288-97.
46. Sonoda, Y., et al., *Akt pathway activation converts anaplastic astrocytoma to glioblastoma multiforme in a human astrocyte model of glioma.* Cancer Res, 2001. **61**(18): p. 6674-8.
47. Won, J., et al., *Membrane frizzled-related protein is necessary for the normal development and maintenance of photoreceptor outer segments.* Vis Neurosci, 2008. **25**(4): p. 563-74.
48. Katoh, M., *Molecular cloning and characterization of MFRP, a novel gene encoding a membrane-type Frizzled-related protein.* Biochem Biophys Res Commun, 2001. **282**(1): p. 116-23.
49. Slaby, O., R. Laga, and O. Sedlacek, *Therapeutic targeting of non-coding RNAs in cancer.* Biochem J, 2017. **474**(24): p. 4219-4251.
50. Saleembhasha, A. and S. Mishra, *Long non-coding RNAs as pan-cancer master gene regulators of associated protein-coding genes: a systems biology approach.* PeerJ, 2019. **7**: p. e6388.
51. Li, L., et al., *HOX cluster-embedded antisense long non-coding RNAs in lung cancer.* Cancer Lett, 2019. **450**: p. 14-21.
52. Jiang, D., et al., *Long Chain Non-Coding RNA (lncRNA) HOTAIR Knockdown Increases miR-454-3p to Suppress Gastric Cancer Growth by Targeting STAT3/Cyclin D1.* Med Sci Monit, 2019. **25**: p. 1537-1548.
53. Team, R., *RStudio: Integrated Development for R.* 2020.
54. Team., R.D.C. *R: A Language and Environment for Statistical Computing.* 2012; Available from: http://www.r-project.org/.