AWARD NUMBER: W81XWH-16-1-0390

TITLE: Role of a Novel Family of Short RNAs, tRFs, in Prostate Cancer

PRINCIPAL INVESTIGATOR: MANJARI KIRAN

CONTRACTING ORGANIZATION: RECTOR & VISITORS OF THE UNIVERSITY
CHARLOTTESVILLE VA 22903

REPORT DATE: November 2018

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
November 2018	Final	1 Aug 2016 – 31 Jul 2018

4. TITLE AND SUBTITLE

Role of a Novel Family of Short RNAs, tRFs, in Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-16-1-0390

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)

Manjari Kiran

E-Mail: mk9ua@virginia.edu

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

UNIVERSITY OF VIRGINIA, N EMMET
STREET, CHARLOTTESVILLE VA
22903-4833

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

tRNA derived fragments, tRFs are a novel class of small RNA which are shown to function by associating to Argonaute proteins. Unlike microRNA, the biogenesis of tRFs is independent of Dicer and Drosha enzymes. Upon meta-analysis of more than 50 small RNA-seq data of different species and cell lines we have previously shown the presence of fragments generated from both 5' and 3' end of the tRNA and called them as tRF-5 and tRF-3, respectively. We are now interested in mining TCGA prostate cancer patients data to identify tRFs and predict potential targets involved in prostate cancer cell migration and proliferation.

15. SUBJECT TERMS
tRF; tRNA-related fragments; Prostate Cancer; Biomarker

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	139	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Page

1. INTRODUCTION:

We have discovered a new class of small RNAs derived from tRNAs known as tRNA derived fragments or tRFs (PMID:19933153). A meta-analysis of more than 50 small RNA datasets led us to decipher that tRFs, although not produced by canonical microRNA producing enzymes like Dicer or Drosha, associate with Argonaute proteins (Ago) and interact with target RNAs based on seed complementarity (PMID: 25270025, PMID: 25392422). In 2016, we showed experimentally that some of these tRFs could guide Ago to regulate gene expression (PMID: 29844106). In the current project, I plan to elucidate the potential role of tRNA-derived fragments as prostate cancer biomarker. Discovering a new biomarker for prostate cancer is significant because early detection and accurate prognosis is very important to cure the disease without over treating many patients who do not have life-threatening condition. Last year, I identified differentially expressed tRFs by comparing the number of distinct types of tRFs in normal and tumor samples of 50 patients. Interestingly, the number of distinct tRFs and average expression of tRFs are higher in tumor compared to normal samples. With the help of my colleague Dr. Canan Kuscu, we mutated the target site on the luciferase reporter and found that mutations that disrupted the pairing of the target with 5' seed of tRFs failed to repress the target presumably by affecting the pairing between tRF and its target. This year, I have used cox-regression on the expression profile of tRFs in different patients to identify prognostic tRFs in prostate cancer.

2. KEYWORDS: tRF; tRNA-related fragments; Prostate Cancer; Biomarker

3. ACCOMPLISHMENTS:

What were the major goals of the project?

Major Task 1: Mining TCGA short RNA raw sequencing data to identify different types of tRNA-derived fragments. (1-6 months)-100% completed
Major Task 2: Predict the targets of tRFs based on sequence similarity. (7-11 months) - 100% completed
Major Task 3: Analyze the expression profile of tRFs in prostate cancer and elucidate the prognostic role of tRFs. (12-15 months) – 100% completed
Major Task 4: Select tRFs with significant predicted targets involved in cell proliferation, cell migration invasion, angiogenesis and experimentally validate whether up or down regulation of tRFs affects their target gene. (16-24 months) – 60% completed

What was accomplished under these goals?

Major Task 1: Mining TCGA short RNA raw sequencing data to identify different types of tRNA-derived fragments. (1-6 months)-100% completed

The steps involved in TCGA data mining are shown in Figure 1. First, I downloaded all the aligned reads for RNA-Seq performed by miRNA-seq experimental strategy for prostate cancer. There are 551 bam files corresponding to 494 prostate cancer patients.

Out of 494 patients, 484 patients are alive and 10 are dead. The paired normal-tumor data is available for 50 patients. There are two patients TCGA-HC-7740 and TCGA-HC-8258 for which three samples are available: 2 corresponding to tumor (01A and 01B) and 1 normal (11A). There is only one patient 'TCGA-V1-A905' with metastatic tumor and the remaining 441 patients have primary tumor. I performed data processing and tRF identification for all the files, from which I am only reporting the results obtained by comparing 50 paired normal-tumor patients. The reads available from TCGA were already trimmed for adapters and mapped against GRCh37 reference genome using BWA-MEM aligners (parameters: samse –n 10) by Marco Marra group from University of British Columbia (Chu et al. 2016). The mapped bam files were then converted to fastq files using bedtools utility with default settings. In order to work with only high quality reads we discarded reads with <30 phred score in 90% of the read length. Now, in the next step, reads were mapped to human tRNA gene to get tRF specific for each patient sample.

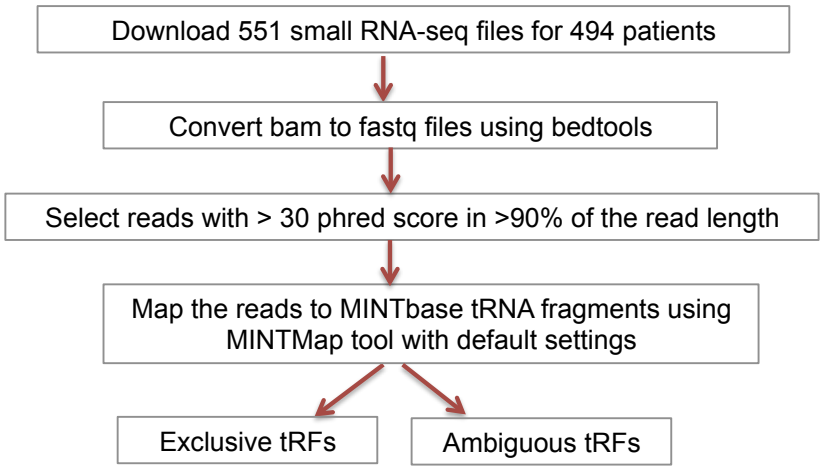


Figure1: Flowchart showing the steps involved in download and processing of data

In order to work with only true positives, I chose a cut-off of 20 RPM and counted combined number of unique tRFs identified by both exclusive and ambiguous method for each patient sample. Around 35 patients have less than 50 tRFs and 25 patients have more than 100 tRFs identified. There are more unique types of tRFs in tumor sample of the patients compared to their normal counterpart (**Figure 2A**).

There can be two possibilities explaining this difference: 1) The parent tRNA of the tRFs are more abundantly expressed in tumor than normal. 2) Some unknown factors are more involved in tRF cleavage in tumor samples or more involved in protection in normal samples of the patients. These two possibilities are not mutually exclusive. To check these possibilities, we can compare the abundance of tRFs grouped based of their parent tRNA isoacceptor.

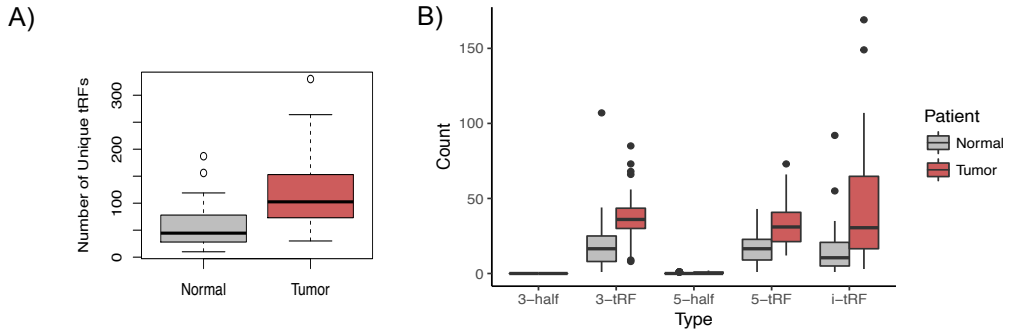


Figure 2: Boxplots showing number of unique tRFs in 50 normal versus 50 tumor patients' samples A) and in different sub-types of tRFs B).

Our group as well as other groups in the field has divided tRNA derived fragments into 5 structural categories.

i) 5-half: longer fragments (>34 nt) that arise from the mature tRNA through cleavage at anticodon of tRNA
ii) 3-half: longer fragments (>34 nt) that are reminder of the mature tRNA following cleavage at anticodon of tRNA
iii) tRF-5/5-tRF: fragments derived after cleavage of mature tRNA at D-loop or the anticodon stem
iv) tRF-3/3-tRF: fragments derived after cleavage of mature tRNA at T-loop or the anticodon stem
v) i-tRF: also known as internal tRFs that can be generated from any other internal sites of tRNA .

I compared the number of distinct types of tRFs in normal and tumor samples of 50 patients. Interestingly, the number of distinct 3-tRF, 5-tRF and i-tRF is significantly higher in tumor than in normal paired samples (P value ~ 2.542e-05) (**Figure 2B**). In contrast, there are no halves identified in either normal or tumor sample. This could be because of running deep-sequencing PCR for only 30 cycles in short RNA-seq library preparation and because of size selection for microRNA sized RNA.
I also noticed higher average expression of tRFs in tumor compared to normal samples (P value = 0.000246) (**Figure 3A**). This again could be because of higher expression or more cleavage of the parent tRNA in tumor than in normal. Among, different structural categories of tRFs, 3-tRFs are the most significantly up-regulated in tumor versus normal (P value ~ 1.387e-05) (**Figure 3B**), which suggests that the cleavage at T-loop is more prominent in tumor samples than normal in prostate cancer patients.

My next aim was to find the top most differentially expressed 3-tRFs in tumor versus normal samples. I first filtered out all the 3-tRFs, with mean expression of less than 20 RPM in 50 tumor patients. There were only 63 3-tRFs which met this criteria. Most of these tRFs are 18 bases long that are annotated as tRF-3a in tRFDB. I found 61 3-tRFs which have significantly higher expression in tumors compared to normal. Interestingly, the top-most differentially expressed 3-tRFs are mostly 24 nucleotides long.

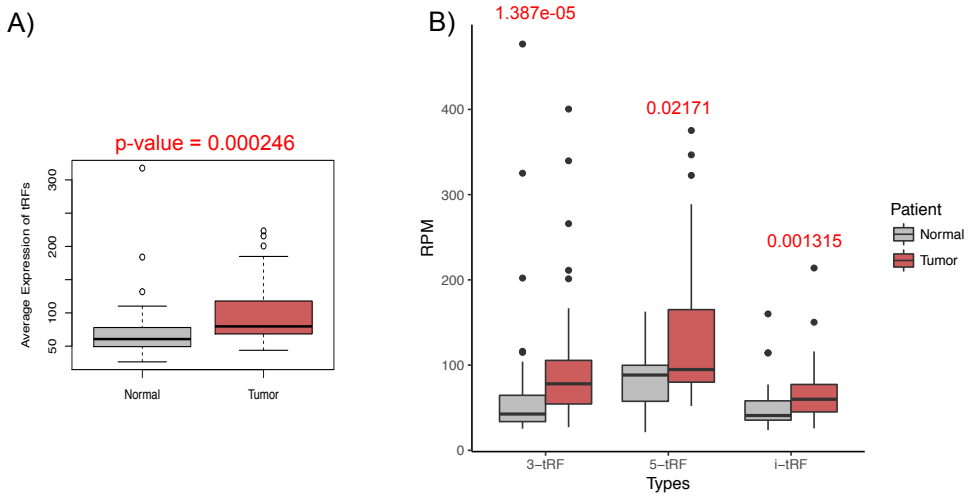
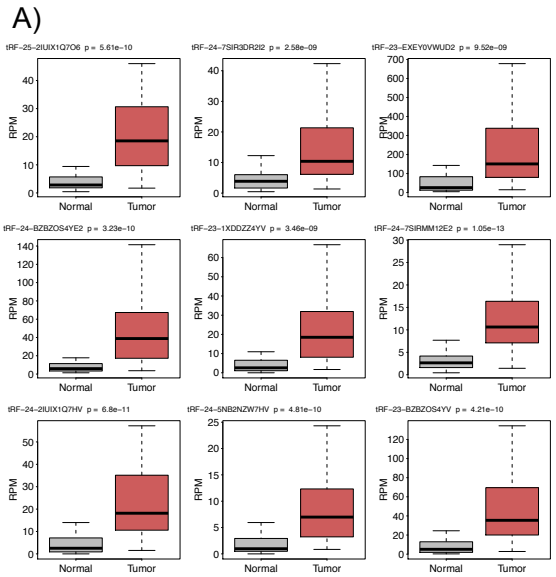


Figure 3: Boxplot showing distribution of average expression of tRFs in 50 normal versus 50 tumor patients' A) and in different sub-types of tRFs B).



B)

tRF Name	tRF sequence	Parent tRNA
tRF-24-7SIRMM12E2	GTTCGATTCCCGGTCAGGGAACCA	GluCTC, GluTTC, PheGAA
tRF-24-2IUIX1Q7HV	CAACTTAACTTGACCGCTCTGACC	ValTAC_MT
tRF-24-BZBZOS4YE2	AACTTAACTTGACCGCTCTGACCA	ValTAC_MT
tRF-23-BZBZOS4YV	AACTTAACTTGACCGCTCTGACC	ValTAC_MT
tRF-24-5NB2NZW7HV	GAGTTAAAGACTTTTTCTCTGACC	ValTAC_MT
tRF-25-2IUIX1Q7O6	CAACTTAACTTGACCGCTCTGACCA	ValTAC_MT
tRF-24-7SIR3DR2I2	GTTCGATTCCCCGACGGGGAGCCA	ProTGG_MT
tRF-23-1XDDZZ4YV	AGTTAAAGACTTTTTCTCTGACC	ValTAC_MT
tRF-23-EXEY0VWUD2	ACTTAACTTGACCGCTCTGACCA	AspGTC

Figure 4: Boxplot showing the distribution of expression level of 9 top-most 3-tRFs obtained by performing Wilcox test which was used to compare mean between normal and tumor prostate cancer patients samples.

Strikingly, more than 70% of 3-tRFs are product of mitochondrial tRNA. 27 and 15 out of 61 differentially expressed 3-tRF are mapping to genomic location of trnaMT_ValTAC_MT_+_1602_1670 and trnaMT_ThrTGT_MT_+_15888_15953, respectively. Further investigation is required to explain this result.

Major Task 2: Predict the targets of tRFs based on sequence similarity. (7-11 months) - 100% completed

In order to decipher how these fragments actually function, I predicted the targets of top-most differentially expressed 3-tRFs based on sequence complementarity. In our previous study, we have also reported numerous tRF-mRNA chimeras based on CLASH (cross-linking, ligation, and sequencing of hybrids) data analysis, which suggested sequence specific interaction of tRFs with RNAs in the cell in Argonaute containing complexes. With the help of my colleague Dr. Canan Kuscu who is one of the primary experimental persons involved in tRF project in the lab, we mutated the target site on the luciferase reporter three bases at a time. We found that mutations that disrupted the pairing of the target with 5' seed of tRFs failed to repress the target. Mutation M3 and M4 in 2-7 nt region from 5'of tRF disrupted repression the most, presumably by affecting the pairing between tRF and its target. We performed this experiment with multiple other tRFs and found consistent results (**Figure 5**).

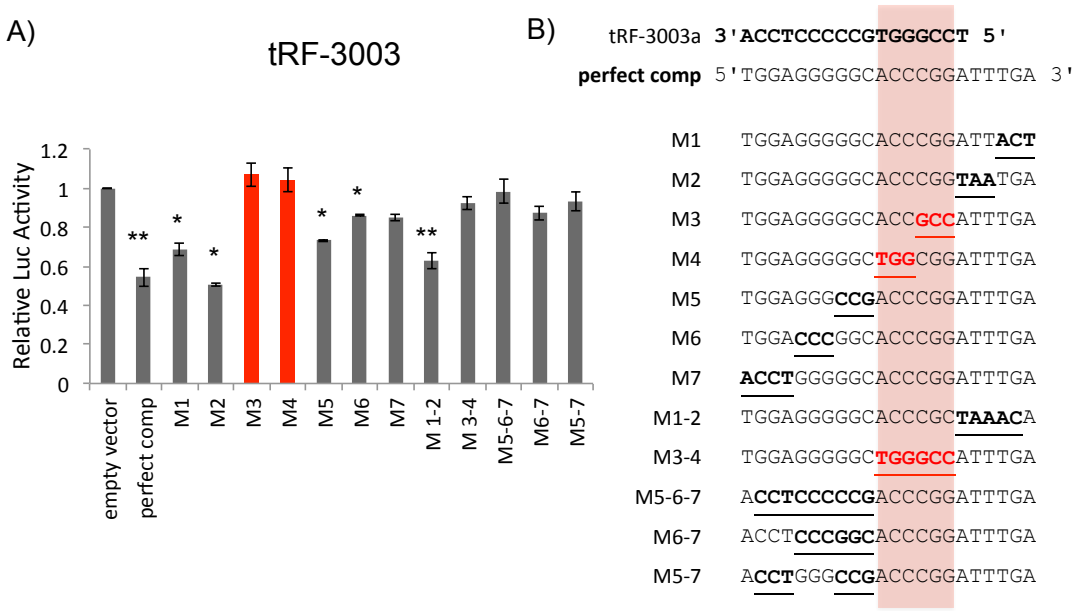


Figure 5: Identification of seed sequence required for target repression by tRFs. A) Luciferase reporter assays with mutant target site at the luciferase reporter upon tRF-3003 overexpression. B) Seed region on tRF-3003 is highlighted in red.

This result suggested that tRFs interact with their targets using their seed sequence similar to miRNA. A script in perl was written to predict targets of the top-most differentially expressed 3-tRFs. The 3'UTR sequence of all RefSeq genes of hg38 genome was downloaded using UCSC Table Browser. In order to remove the bias caused by genes with many isoforms, I considered only the most highly expressed isoform for a gene in Hela cells as identified by 3p-seq by Bartel group in 2014 (Nam et al. 2014). A total of 9294 sequences were examined for the complementarity of various seed

sequences. Considering that a tRF interacts with its target using seedmer similar to miRNA, each 3UTR sequence was first scanned for 8mer followed by 7mer-m8, followed by 7mer-A1 and the remaining pool was scanned for 6mer.

In total, I found 2977 targets for tRF-24-2IUIX1Q7HV and 2257 targets for tRF-23-EXEY0VWUD2, the two tRFs identified from previous step as the most differentially expressed in tumor versus normal samples. As expected, due to the difference in seed length and therefore probability to find matching sequence, most of the predicted targets identified belong to 6mer category and least belong to 8mer category. My next aim is to find miRNA and RNA binding proteins as potential targets of these tRFs.

Major Task 3: Analyze the expression profile of tRFs in prostate cancer and elucidate the prognostic role of tRFs. (12-15 months) – 100% completed

Last year, I downloaded all the aligned reads for RNA-Seq performed by miRNA-seq experimental strategy for prostate cancer from TCGA gdc portal. In total, I identified 44,665 unique tRFs for 494 prostate cancer patients using MINTmap mapping tool (PMID: 28220888). I hypothesize that there will be differences in tRF production depending on the transformation and growth state of prostate cells. So, one of the major goal this year was to identify prognostic tRF for prostate cancer survival. Out of 44,665 unique tRFs identified in 494 patients only 554 tRFs were expressed more than median 1RPM (Reads per Million) in all patients. I considered two time points for survival analysis: Overall Survival (OS) and Progression free survival (PFS). The median overall survival of prostate cancer patients in TCGA dataset is 924 days with 484 alive and 10 dead patients. The median progression free survival for prostate cancer patient in TCGA dataset is 788 days. I calculated cox-coefficient (a measure of association with survival) for each tRF in order to identify tRFs associated with prostate cancer patients' survival. To achieve this, I used Cox proportions hazards model, which is a standard regression method for studying survival data. I found three tRFs with FDR cut-off of less than 0.20 important for prostate cancer progression. tRF-16-9LON4VD and tRF-23-94U47P2904 are internal tRFs associated with overall survival of prostate cancer patient whereas tRF-21-WE884U1D is a tRF from 3' end of tRNA associated with progression free survival. Table1 shows tRF sequence, length, summary and statistics obtained by cox-regression for all the three tRFs.

Table1 : Summary of prognostic tRFs for prostate cancer survival identified by cox proportional hazards model.

tRFs	Summary	Sequence	Interval	HR	P-val (FDR)
tRF-21-9O9NF5W8B	i-tRF (GlnTTG and GlnCTG) 21 nt, 13th base from 5'end	TGGTGTAATGGTTAGCACTCTGG	OS	0.15	0.00064 (0.176554)
tRF-23-94U47P2904	i-tRF GlnTTG and GlnCTG), 23 nt, 9th nt from 5' end	TGTAATGGTTAGCACTCTGGA	OS	0.04	0.00044 (0.176554)
tRF-21-WE884U1DD	**3'-tRF TyrGTA), 21 nt, till CC 3' end**	**TCGATTCCGGCTCGAAGGACC**	**PFS**	**-0.27**	**0.000113 (0.0620)**

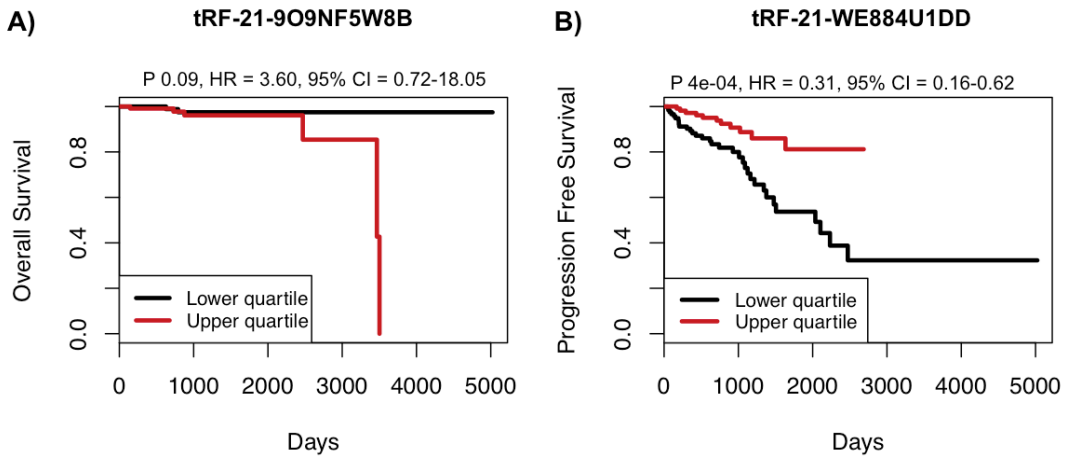


Figure 6: Kaplan-meier plots showing prognostic tRFs based on overall survival A) tRF-21-9O9NF5W8B and based on progression free interval B) tRF-21-WE884U1DD. The black line corresponds to patients with low expression (below 25th percentile) and red line corresponds to patients with high expression (above 75th percentile) for both tRFs in respective plots.

tRF-21-9O9NF5W8B is associated with poor prognosis whereas tRF-21-WE884U1DD is associated with good prognosis of the prostate cancer (**Figure 6**). The overall survival of patients with low expression was not statistically different from patients with high expression of tRF-23-94U47P2904. Since, the result obtained for tRF-21-WE884U1DD was most significant, I decided to follow tRF-21-WE884U1DD for further analysis. This tRF is annotated as tRF-3030b in tRFdb. For simplicity, tRF-21-WE884U1DD will be referred as tRF-3030b throughout the report. In order to predict function of tRF-3030b, I applied an approach where I calculated the fold change difference of mRNAs in patients with high expression of tRF-3030b (more than 75th percentile) to the patients with low expression of tRF-3030b (less than 25th percentile). The log-fold-change of genes thus obtained was used for gene set enrichment analysis. There were 9 pathways enriched in up-regulated and 41 pathways enriched in down-regulated genes in the patients with high versus low expression of tRF-3030b. Interestingly, among the mRNAs that are down regulated in high versus low tRF patients are genes involved in immune and inflammatory response, genes involved in epithelial to mesenchymal transition and cell-cycle progressions (**Figure 7**). These gene-set enrichments suggest a conventional tumor suppressor phenotype associated with tRF-3030b in prostate cancer.

Major Task 4: Select tRFs with significant predicted targets involved in cell proliferation, cell migration invasion, angiogenesis and experimentally validate whether up or down regulation of tRFs affects their target gene. (16-24 months) – 60% completed

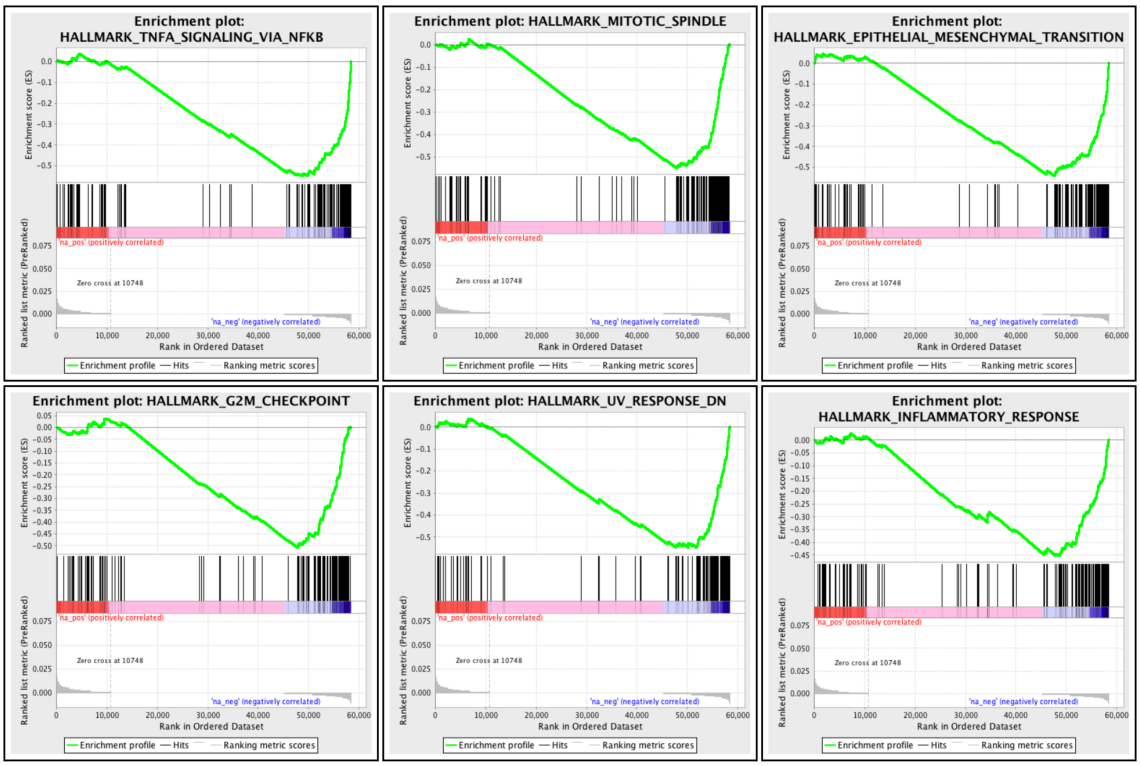


Figure 7: Screenshot of GSEA result obtained for genes in two groups of patients with high versus low level of tRF-21-WE884U1DD.

This analysis suggested the role of tRF-3030b in cell proliferation and epithelial to Mesenchymal transition. To test the effect of decreased levels of endogenous tRFs on cell proliferation, with the help of my colleague Dr. Canan Kuscu we transfected tRF-complementary oligonucleotides to knock down tRFs. tRF-sponging oligonucleotides are synthesized by commercial source (Qiagen) with LNA (lock nucleic acid) modifications that will enhance their stability and specificity. A non-targeting sequence is used as negative control. Upon transfection of various concentrations of tRF-sponging LNAs, cell proliferation is measured indirectly by colorimetric MTT assay in multi-well plate reader. Briefly, 72 hours post LNA transfection, cells were treated with MTT reagent, a yellow tetrazole that will be reduced to purple formazan by alive cells. Purple formazan will then be solubilized in solution to give absorbance at 570nm. Absorbance values from MTT assay will be subtracted to the

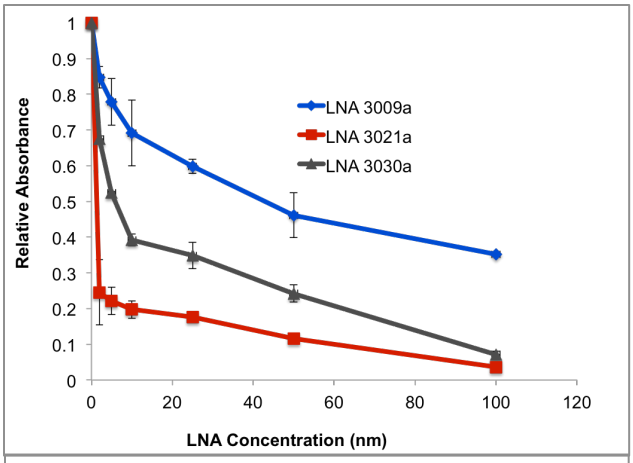


Figure 8: Sponging of endogenous tRFs decrease cell proliferation. tRF-sponging LNAs were transfected for three tRFs – 3309a, 3021a, 3030a.

background reading and normalized to the control wells. I am now repeating this experiment for tRF-3030b in prostate cancer cell lines like LnCap and PC3 cultured in the lab.

What opportunities for training and professional development have the project provided?

In two years, this project has helped me to exploit my existing computational skill and also has helped me to endeavor many new professional and technical abilities required to be an independent researcher. I worked on several related projects, which led to multiple publications either as first author (PMID: 30392137, PMID: 30037979) or in collaboration (PMID: 29991527, PMID: 29844106, PMID: 27906128). These led me to secure an Assistant professor position in a reputed University in India. I will be extending my current research as an independent researcher hereafter.

How were the results disseminated to communities of interest?

"Nothing to Report."

What do you plan to do during the next reporting period to accomplish the goals?

"Nothing to Report."

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

The aim of the project is to identify a specific biomarker for better prognosis of prostate cancer. Small RNAs like microRNAs have been linked to prostate cancer pathogenesis. Last year, after mining small RNA data available for prostate cancer patient at TCGA, I found many tRFs overexpressed in tumor compared to normal tissue. The results obtained supported the existence of an entirely new group of molecular drivers of prostate cancer. This year, I have identified tRFs that can be used for predicting the survival of prostate cancer patient. Several analyses performed as a part of this project indicate that these prognostic tRFs are involved in cell proliferation and tumor progression. These tRFs could serve as biomarker for early cancer detection or prognosis.

What was the impact on other disciplines?

“Nothing to Report.”

What was the impact on technology transfer?

“Nothing to Report.”

What was the impact on society beyond science and technology?

"Nothing to Report."

5. CHANGES/PROBLEMS: "Nothing to report"

6. PRODUCTS:
Journal publication:

C Kuscu[1], P Kumar[1], **M Kiran[1]**, Z Su[1], A Malik[1], A Dutta[1]**.** tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. RNA 24 (8), 1093-1105. 2018
[1]Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA, USA

M Kiran[1], A Chatrath[1], X Tang[2], DM Keenan[2], A Dutta[1]. A Prognostic Signature for Lower Grade Gliomas Based on Expression of Long Non-Coding RNAs. Molecular Neurobiology, 1-13. 2018
[1]Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA, USA
[2]Department of Statistics, University of Virgini, Charlottesville, VA, USA

MA Cichewicz[1], **M Kiran[1]**, RK Przanowska[1], E Sobierajska[1], Y Shibata[1], A Dutta[1]. MUNC, an Enhancer RNA Upstream from the MYOD Gene, Induces a Subgroup of Myogenic Transcripts in trans Independently of MyoD. Molecular and cellular biology 38 (20), e00655-17. 2018
[1]Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA, USA

BJ Reon[1], BTR Karia[1], **M Kiran[1]**, A Dutta[1]. LINC00152 Promotes Invasion through a 3′-Hairpin Structure and Associates with Prognosis in Glioblastoma. Molecular Cancer Research 16 (10), 1470-1482. 2018
[1]Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA, USA

Note: Department of Defense fellowship is fully acknowledged in the all the manuscript published by PI and is attached in appendix.

Books or other non-periodical, one-time publications. "Nothing to report"

Other publications, conference papers, and presentations:
Dutta A, Kumar P, **Kiran M**, Kuscu C. Transfer RNA Fragments (tRFs): a Novel Class of Non-micro Short RNAs that Uses Ago1, 3 and 4 to Repress Specific Target RNAs Through 5′ Seed Sequences.The FASEB Journal 30 (1 Supplement), 1054.5-1054.5
(This abstract is from the Experimental Biology 2016 Meeting)

Website(s) or other Internet site(s) "Nothing to report"

Technologies or techniques "Nothing to report"

Inventions, patent applications, and/or licenses "Nothing to report"

Other Products "Nothing to report"

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name: Manjari Kiran "no change"

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

"Nothing to report."

What other organizations were involved as partners?

"Nothing to report."

8. SPECIAL REPORTING REQUIREMENTS None

9. APPENDICES:

tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer independent manner

Canan Kuscu[1, 2], Pankaj Kumar[1, 2], Manjari Kiran[1], Zhangli Su[1], Asrar Malik[1], Anindya Dutta[1,*]

[1]Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, Virginia, USA 22901

[2] Co-author

* Correspondence: Anindya Dutta

Email: ad8q@virginia.edu; Tel: [+1-434 - 924 – 2466]; Fax: [+1-434 - 924 –5069]

Running title: tRNA fragments regulate genes post-transcriptionally

Key words: tRNA fragments, tRF, post-transcriptional gene regulation, small non-coding RNA

Abstract:

tRNA related RNA fragments (tRFs), also known as tRNA derived RNAs (tdRNAs), are abundant small RNAs reported to be associated with Argonaute proteins, yet their function is unclear. We show that endogenous 18 nucleotide tRFs derived from the 3' ends of tRNAs (tRF-3) post-transcriptionally repress genes in HEK293T cells in culture. tRF-3 levels increase upon parental tRNA over-expression. This represses target genes with a sequence complementary to the tRF-3 in the 3' UTR. The tRF-3-mediated repression is Dicer-independent, Argonaute-dependent and the targets are recognized by sequence complementarity. Furthermore, tRF-3:target mRNA pairs in the RNA Induced Silencing Complex associate with GW182 proteins, known to repress translation and promote the degradation of target mRNAs. RNA-seq demonstrates that endogenous target genes are specifically decreased upon tRF-3 induction. Therefore, Dicer-independent tRF-3s, generated upon tRNA overexpression, repress genes post-transcriptionally through an Argonaute-GW182 containing RISC via sequence matches with target mRNAs.

37 **Introduction:**

38 There are many different classes of cellular small RNAs, including microRNAs (miRNAs), Piwi-
39 interacting RNAs (piRNAs) and tRNA related fragments (tRFs). tRFs are 14-32 base-long RNAs
40 derived from tRNAs that have been identified from bacteria to humans with high abundance. The <26 nt
41 long tRFs can be classified into four main groups: tRF-5s and tRF-3s, from the extreme 5' and 3' ends
42 of mature tRNAs respectively; i-tRFs, internal fragments spanning anywhere in the mature tRNA but
43 not mapping to the extreme 5' or 3' ends of mature tRNA; and tRF-1s, from the 3' trailer of precursor
44 tRNA ((Lee et al. 2009; Haussecker et al. 2010; Telonis et al. 2015); reviewed in (Keam and Hutvagner
45 2015; Kumar et al. 2016)). Several groups have developed bioinformatics tools to analyze tRNA derived
46 fragments from small RNA sequencing and generated several databases showing differential expression
47 of tRFs in cell-lines, tissues and disease states (Kumar et al. 2015; Selitsky and Sethupathy 2015; Zheng
48 et al. 2016; Pliatsika et al. 2018; Thompson et al. 2018).

49 The longer, 30-36 base long fragments are also called tiRNAs/tRNA halves (tRHs) and have
50 been tied to stress response (reviewed in (Saikia and Hatzoglou 2015)). These 5'tiRs/tRHs, produced by
51 cleavage in the anti-codon loop, have essential roles in transgenerational gene regulation by metabolic
52 stress: mice with high fat diet have higher levels of 5'tIRs/tRHs in their sperm and these serve as
53 signaling molecules in their offspring (Chen et al. 2016; Sharma et al. 2016).

54 As for tRF-3s, <26 base fragments from the 3'end of mature tRNAs, several studies have
55 suggested more diverse functions. Recently, Schorn et al. studied tRF-3s in mouse stem cells by
56 comparing wild type cells with *Setdb1*-/- cells lacking methylation on Histone 3 Lysine 9 (H3K9me3).
57 They observed elevation of 18nt tRF-3a molecules and showed that they are necessary for the repression
58 of LTR-retrotransposons (Schorn et al. 2017). Kim et al. identified a tRF-3 from LeuCAG3 tRNA (tRF-
59 3011b), which they called LeuCAG3'tsRNA. This tRF is important for cell proliferation and enhances

60 ribosome biogenesis by binding at least two ribosomal protein mRNAs, RPS28 and RPS15. They also
61 showed that tRF-3011b is upregulated in hepatocellular carcinoma and knock down of tRF-3011b
62 inhibits tumor growth (Kim et al. 2017). In addition, Maute et al. showed that a tRNA derived miRNA
63 (CU1276: tRF-3027b) is down regulated in B-cell lymphoma. tRF-3027b, 22nt in length, physically
64 associates with Argounate proteins and represses the expression of single strand DNA binding protein
65 RPA1 in a sequence-dependent manner. Upregulation of RPA1, due to down regulation of tRF-3027b in
66 Burkitt lymphoma cell lines, results in aberrant proliferation of cells (Maute et al. 2013).

67 In addition to functions listed above, a meta-analysis of available short RNA sequencing data
68 suggested that tRFs are present in mouse embryonic stem cells that are *Dicer*-/- (Kumar et al. 2014).
69 Moreover, analysis of PAR-CLIP data (Hafner et al. 2010) identified abundant tRF reads in Argonaute
70 complexes, the main effector proteins in the miRNA pathway (Kumar et al. 2014). This analysis
71 indicated that some tRF-3s and tRF-5s could bind to Ago proteins in a manner similar to miRNAs and
72 use their 5' seed sequence of 7-8 bases to bring their target mRNAs into the Ago complexes (Kumar et
73 al. 2014). Most intriguingly, analysis of CLASH data from (Helwak et al. 2013) identified more tRF-3-
74 mRNA chimeras than miRNA-mRNA chimeras associated with Ago1 (Kumar et al. 2014). All these
75 lines of evidence suggested a potential function of tRF-3s which interact with Argonaute proteins in
76 regulating gene expression using a mechanism similar to miRNAs.

77 In this study, we focused on 18-nt tRF-3s which interact with Argonaute complexes. To
78 experimentally test whether tRF-3s actually repress gene expression, we devised ways to up-regulate
79 endogenous tRF-3 levels generated from tRNAs. We focused on three tRFs that were reported to be
80 associated with Argonaute proteins in our previous report (Kumar et al. 2014). Overexpression of a
81 particular tRNA (tRNA LeuAAG, tRNA CysGCA or tRNA LeuTAA) resulted in production of the
82 specific tRF-3 (tRF-3001, tRF-3003 and tRF-3009, respectively) (nomenclature based on tRFdb (Kumar

83 et al. 2015)). The tRF-3s specifically downregulate expression of luciferase reporters with a
84 complementary sequence in the 3' UTR. This repression is Dicer-independent but Argonaute dependent,
85 and also dependent on match to the seed sequence at the 5' end of the tRF. tRF-3:mRNA pairs associate
86 with the "effector" GW182/TNRC6 proteins in the RISC, known to promote translational repression and
87 target mRNA degradation. Furthermore, RNA-seq of HEK293T cells overexpressing the tRF-3009a
88 showed that mRNA targets of tRF-3009a are significantly decreased, consistent with their degradation
89 by the tRF-3s. Thus, tRF-3s, generated independent of miRNA biogenesis proteins such as Dicer and
90 Drosha, are able to enter into Argonaute-GW182 containing RISC and regulate mRNAs post-
91 transcriptionally via sequence complementarity.
92

93 **Results:**

94 **Overexpression of tRNA results in production of the corresponding tRF-3**

95 tRF-3s are 18-22 nucleotide small RNAs of two distinct sizes; tRF-3a (~18 nt) and tRF-3b (~22
96 nt) (Kumar et al. 2014) (Figure 1A). We focused on three tRF-3a that we knew to associate with AGO
97 proteins in PAR-CLIP and CLASH data: tRF-3001a, tRF-3003a and tRF-3009a (Kumar et al. 2014).
98 These three tRF-3s have very distinct lengths and locations on the parental tRNAs and are usually the
99 most abundant fragments from the tRNA (Figure 1D and Supplementary Figure S1). They have been
100 also detected in several different small RNA sequencing datasets (see tRFdb:
101 http://genome.bioch.virginia.edu/trfdb/). We are aware of two technical issues that may limit the
102 acquisition of a full repertoire of tRFs by small RNA sequencing. First, the many modifications of
103 tRNAs or tRFs could block the reverse transcriptase producing smaller cDNA products than the actual
104 RNA molecules. Second the size selection in many small RNA sequencing protocols may intentionally
105 exclude small RNAs below or above a particular size. Despite these challenges, it is remarkable that
106 small RNA sequence libraries from tens of laboratories around the lab yield tRFs of consistent size,
107 sequence and abundance.

108 We overexpressed the parental tRNAs in HEK293T cells and assessed tRNA and tRF levels by
109 qRT-PCR. The overexpression of tRNA LeuAAG, tRNA CysGCA and tRNA LeuTAA resulted in
110 overproduction of the corresponding tRF-3s by 7-70 fold relative to endogenous tRF levels (Figure 2A
111 and 2B). Since PCR based detection of tRF-3s relies on reverse transcription (which could produce
112 specific short cDNAs due to polymerase block by specific tRNA modifications), we also performed
113 northern blot analysis to confirm the induction of tRF-3a and -3b in all three cases (Figure 2C). To prove
114 that tRF-3s produced by tRNA overexpression can be loaded into Argonaute, we performed Argonaute

115 immunoprecipitation followed by northern blotting to show that tRF-3009a but not tRF-3009b is loaded
116 into Argonaute (Figure 2D and Supplementary Figure S2).
117
118 **tRF-3s silence gene expression through sequence complementarity.**
119 Luciferase reporters with perfect complementary sequences to the tRFs in the 3' UTR were co-
120 transfected with the tRNA. When tRF-3001 is expressed by overexpressing tRNA-LeuAAG and the
121 luciferase 3'UTR has a perfect complementary sequence to tRF-3001, the luciferase activity is repressed
122 to 40% (Figure 3A). Luciferase reporters with complementarity to the cognate tRFs were down
123 regulated similarly upon tRF-3003 and tRF-3009 overexpression (Figure 3A).
124 Titration of the amount of tRNA expressing plasmid showed that the downregulation of
125 luciferase is directly correlated with the amount of the tRNA expressing plasmid (Figure 3B and
126 Supplementary Figure S3). We tested the specificity of the downregulation by transfecting tRNA
127 expressing plasmid and luciferase reporter plasmid with complementarity to the other two tRF-3s and no
128 repression was observed (Figure 3C). For example tRF-3001 producing plasmid selectively repressed
129 the luciferase reporter with complementarity to tRF-3001, but not those with complementarity to tRF-
130 3003 or tRF-3009 and so on (Figure 3C). To confirm the repression observed in luciferase activity is
131 mediated by tRF-3a and not due to other forms of tRFs, we transfected an 18-nt tRF-3009a mimic and
132 observed similar repression of a luciferase reporter with the "perfect comp" target site (Supplementary
133 Figure S4).
134
135 **tRF-3s repress targets with seed sequence matches**
136 Having identified the gene repression activity of tRF-3s, we wanted to characterize the base
137 pairing essential for target repression. We mutated the target site (tRF-complementary sequence) on the

138 luciferase reporter to check whether disrupting certain base pairs between the tRF and the luciferase
139 target would abolish luciferase repression. Note that microRNAs often recognize their target mRNAs by
140 complementarity to a so called seed sequence at the 5' end, which usually uses an A:T pair at the
141 extreme 5' end followed by 6-7 bases. In other words the mRNA target often pairs with the bases 1-7
142 or bases 2-8 at the 5' end of the microRNA.

143 As mentioned above, tRF-3s can have two distinct lengths: 18nt (tRF-3a) or 22nt (tRF-3b), and
144 the short RNA sequence data analysis suggests that tRF-3001 only has "a" form, while tRF-3003 and
145 tRF-3009 are present in both "a" and "b" forms (Kumar et al. 2015). The luciferase reporters have
146 perfect complementarity to the longer tRF-3b so that both forms of tRFs (in Figure 4B and C), if
147 expressed, could repress the target. Mutations that disrupted pairing of the target with the 5' seed of each
148 tRF-3a diminishes repression (Figure 4A: M1, M2, M3; Figure 4B: M3, M4; Figure 4C: M3, M4). The
149 mutational analysis also showed that repression was mostly mediated via tRF-3a, because mutations at
150 the extreme 3' end of the target, that would disrupt pairing with the seed sequence of the longer tRF-3b,
151 but not tRF-3a, continued to be repressed by the tRNA overexpression at almost the same level as the
152 perfectly complementary target (Figure 4B and C: M1, M2).

153 Perfect complementarity to the entire length of the tRF was not required for repression. For
154 example, tRF-3003a can still repress a target that has mutations at the middle of the tRF-mRNA pair
155 indicating tolerance for a bulge structure in the middle of the tRF-mRNA pairing (Figure 4B: M5).
156 Mutations further away from canonical seed regions also allowed repression (Figure 4A, 4B: M6). In all
157 cases, however, mutations outside of canonical seed regions (Figure 4A: M5; Figure 4B and C: M5, M6
158 and M7) showed less repression than with the perfect match target, suggesting that tRFs appear to
159 require additional complementarity outside the seed sequence for maximal repression. In parallel, we
160 also performed luciferase assay with the reporter mutants using tRF-3009a mimic which is a synthetic

161 siRNA with same sequence of tRF-3009a. The results were similar to what we observed with tRNA
162 overexpression (Supplementary Figure S4) confirming that the effects we have seen are due to tRF-
163 3009a produced from tRNA LeuTAA. Thus, the rules of repression are similar to microRNAs, in that
164 the target should be complementary to the 5' seed sequence of the tRF. Though perfect complementarity
165 to the entire length of the tRF is not essential for repression, additional complementarity downstream
166 from the seed sequence facilitates repression.

167

168 **Gene silencing through tRF-3 is dependent on Argonaute proteins.**

169 The seed sequence properties of tRF-3 (Figure 4) is reminiscent of microRNA rules and
170 consistent with our hypothesis that tRF-3 can enter Ago complexes to perform microRNA-like
171 functions. Our previous analysis of AGO PAR-CLIP showed that tRF-3s associate with Argonaute
172 proteins (Figure 1B and (Kumar et al. 2014)). Moreover, AGO1 CLASH analysis revealed that tRF-3s
173 form chimeras with mRNAs in AGO1 (Figure 1C and (Kumar et al. 2014)). We have also demonstrated
174 by northern blot that tRF-3009a produced from tRNA-LeuTAA overexpression is loaded into Argounate
175 (Figure 2D). To test whether Argonaute proteins (Ago) are essential for the gene repression function of
176 tRF-3s, we performed the luciferase reporter assays upon Ago-1, -2 and -3 knock down by siRNA.
177 Ago-4 levels also decreased upon knock down of Ago1, 2 and 3 (Supplementary Figure S5A). As seen
178 in Figure 5A, the repression of the luciferase reporter by tRF-3009 was diminished when the Argonaute
179 proteins were downregulated. Repression by tRF-3001 or tRF-3003 was also attenuated in response to
180 lower Argonaute protein levels (Supplementary Figure S5B and S5C).

181 Catalysis by Argonaute may generate Dicer independent miRNA, miR-451 (Cheloufi et al.
182 2010). We therefore checked the levels of tRF-3s in Ago knock down conditions since loss of repression
183 seen in Figure 5A might be due to a defect in the biogenesis of tRF. By northern blot analysis, tRF-3009

184 is still generated as efficiently as in WT cells after Ago knockdown (Figure 5B). Therefore, Ago is
185 necessary for gene repression by tRF-3s but not for their biogenesis.
186
187 **Biogenesis of tRF-3 is independent of Dicer, Drosha and Exportin 5.**
188 One of the major enzymes in miRNA pathway is Dicer, which cuts out the loop in the stem-loop
189 hairpin structure of the precursor miRNA (pre-miRNA) to generate miRNA (Bernstein et al. 2001). The
190 role of Ago in tRF-3 mediated repression prompted us to check whether Dicer is important for tRF-3
191 mediated repression. Dicer could have a role in loading tRF-3 into Ago complexes, or in tRF-3
192 biogenesis. We obtained Dicer knock-out 293T cell lines that have been shown to be defective in
193 repressing microRNA targeted luciferase reporters (Bogerd et al. 2014). Analysis of small RNA
194 sequencing data from these cells showed that although Dicer is required for microRNA biogenesis, it is
195 not required for the biogenesis of tRF-3s including tRF-3001, tRF-3003 or tRF-3009 (Figure 5C and
196 Supplementary Figure S6B). Furthermore, the luciferase repression by tRNA generated tRF-3 is still
197 observed in the *Dicer* knock-out cells, and is even more pronounced than in wild type cells (Figure 5D),
198 most probably because more Argonaute protein becomes available due to the failure to produce cellular
199 miRNA. Therefore, Dicer is not required to produce the 18-nt tRF-3a nor is essential for tRF-3 loading
200 into Ago.
201 Our previous report showing that tRF abundance was unchanged in the absence of Dicer or
202 DGCR8 was derived from embryonic stem cells (Kumar et al. 2014). Besides the Dicer knockout 293T
203 cells described above, *Dicer*-/-, *Drosha*-/- and *Exportin 5*-/- HCT116 colon cancer cells have been
204 generated using CRISPR-Cas9 technology, providing a second opportunity to test the importance of
205 these enzymes in tRF biogenesis (Kim et al. 2016). Analysis of the small RNA sequences from these
206 cells showed that although Dicer and Drosha are important for microRNA levels, they are dispensable

207 for tRF-3 generation (Supplementary Figure S6C and S6D). Exportin 5 depletion did not affect tRF-3
208 levels (Supplementary Figure S6D), similar to previous observations that Exportin 5 is dispensable for
209 miRNA biogenesis (Supplementary Figure S6C) (Kim et al. 2016). tRF-3016 is an exception in that it is
210 significantly repressed in the Exportin 5 deleted cells due to unknown reasons.

211

212 **tRF-3s interact with GW182/TNRC6A, the effector protein for translational repression and**
213 **mRNA degradation.**

214 In addition to Dicer and TRBP, AGO interacts with GW182 proteins in RISC complex. GW182
215 interaction with AGO is necessary for translation repression and degradation of target mRNA (Eulalio et
216 al. 2008). PAR-CLIP of GW182 showed that it interacts with miRNAs and mRNA targets (Hafner et al.
217 2010). To test whether tRF-3s also interact with GW182 proteins, we investigated the association of
218 tRF-3s with GW182 (TNRC6A) and its paralogs (TNRC6B and C) by analyzing PAR-CLIP data from
219 HEK293 cells (Hafner et al. 2010). As seen in Figure 6A, TNRC6A/B/C associate with some tRF-3s,
220 but not all. The most abundant association was observed for TNRC6A, in which case a few tRF-3s are
221 present at levels comparable to miRNA levels (>1000 RPM), although in general tRF-3s associate with
222 TNRC6 proteins at a lower level compared to microRNAs. Moreover, the locations of the T to C
223 mutations in the PAR-CLIP data indicate that tRF-3s directly interact with TNRC6 proteins at position
224 12 and 14, while sparing bases 1-6 that overlap with the seed (Figure 6B). This interaction is reminiscent
225 of miRNA-TNRC6 interaction (Hafner et al. 2010), at position 11 and 13, with the bases 1-6 in the seed
226 sequence being spared from the cross-link.

227 The absence of any cross-link of the TNRC6 with the seed sequence of the tRF-3 or microRNA
228 is likely because the seed is paired with the target RNA even in the TNRC6 complexes. Since
229 GW182/TNRC6A associates with AGO in the RISC, we wondered whether the targets of the tRF-3 may

230 also be detected in the TNRC6 PAR-CLIP data. We therefore searched in the PAR-CLIP data for target
231 RNA reads with complementarity either to seeds of microRNAs, or specifically to tRF-3s but not
232 microRNAs. Targets that were detected were aligned with the T to C mutation marking the cross-link at
233 the center along with 20 bases up and downstream from the cross-link to produce cross-link-centered-
234 RNAs (CCRs). Target RNAs complementary to microRNAs can be detected in the CCRs associated
235 with TNRC6 (Figure 6C). Most interesting, target RNAs with complementarity to tRF-3 seeds (1-7 mer
236 or 2-8 mer) are also associated with the TNRC6 (Figure 6D). In the case of both the microRNAs and the
237 tRF-3s, the most frequent complementarity to the seed is seen just downstream from the cross-link
238 center, similar to what is seen with the microRNA:target or tRF-3:target pairs in Ago1-4 PAR-CLIP data
239 (Hafner et al. 2010). This result is consistent with the proposal that, as with microRNAs, GW182
240 associates with the tRF3:target in the Ago containing RISC to promote translation repression and target
241 mRNA degradation.
242
243 **RNA-seq analysis reveals that tRF-3009a represses endogenous mRNA targets.**
244 GW182 associated with miRNA-mRNA loaded RISC complex mediates the degradation of
245 mRNA targets (reviewed in (Jonas and Izaurralde 2015)). Since tRF-3s paired with their target RNAs
246 are also found associated with GW182 proteins, we first checked the mRNA levels of luciferase in
247 luciferase reporters containing a perfect complementary site to tRF-3009. In this experiment, we assayed
248 both the luciferase protein level by luciferase assay and the luciferase mRNA level by qRT-PCR from
249 the same cells. As seen in Figure 7A, mRNA and protein levels of Renilla luciferase are both
250 downregulated to ~50%. This suggests the tRF-3 mediated repression occurs mostly by degrading the
251 target RNA, most likely by the de-adenylation/de-capping followed by exonuclease digestion that has
252 been proposed for microRNA targets.

253 Next, we analyzed cellular mRNA expression changes by RNA-sequencing after tRNA
254 overexpression mediated tRF overproduction. Overexpression of tRNA Leu TAA, which produces tRF-
255 3009a, repressed target mRNAs with 3'UTRs bearing complementarity to at least the 6-mer seed
256 sequence of the tRF, as predicted by RNA22 (Miranda et al. 2006), relative to RNAs that do not have
257 such complementarity (Figure 7B). This experiment was repeated again and cumulative distribution
258 function plots of the second replicate shows a similar trend (Supplementary Figure S7).

259 The top 5 repressed targets from RNA-seq contain not only base complementarity to the tRF-
260 3009a seed but also have some complementarity downstream of the seed (Figure 7C). We first validated
261 the decrease in RNA of the top 5 repressed targets by qRT-PCR (Figure 7D). Moreover, we cloned the
262 3'UTR of these targets into luciferase reporter plasmids and performed luciferase reporter experiments.
263 Luciferase reporters containing the 3'UTRs of all these genes are repressed upon overexpression of
264 tRNA-LeuTAA producing tRF-3009s (Figure 7E). This result independently validates the repression
265 observed in the RNA-seq experiment.

266 **Discussion:**

267 tRNA derived RNA fragments (tRFs) have been discovered and characterized from small RNA
268 sequencing, and so certain technical limitations should be recognized. First, base modifications in
269 tRNAs and tRFs could block the progression of the reverse-transcriptase, leading to artificially truncated
270 cDNA products. To guard against this, key results in this paper have been validated by Northern
271 blotting for tRFs of the correct size. Second, many of the cloning protocols will not clone fragments that
272 lack 5' phosphate or have a 2'- 3' cyclic phosphate at the end (as with 5' tIRs/tRHs). Last, size selection,
273 often used during microRNA sequencing, can artificially limit the recovery of tRNA fragments below or
274 above a certain size. Because of these limitations, small RNA sequencing data needs to be analyzed
275 while paying attention to these technical differences, as we have done here. In this paper we focus on

276 tRF-3s detected in all small RNA sequencing libraries, present at reasonable abundance relative to other
277 fragments and shown to interact with Argonaute proteins. Undoubtedly new tRFs will emerge and the
278 abundance of tRFs will increase as new techniques are deployed to sequence tRNAs and tRFs that
279 overcome limitations due to tRNA modifications (Cozen et al. 2015; Zheng et al. 2015). In particular,
280 cP-RNA-seq, developed to selectively amplify and sequence RNAs that end in cyclic phosphates, will
281 increase the recovery of 5'tiRs/tRHs (Honda et al. 2016). These new tRFs may have completely
282 different functions and may not act in the manner of the tRF-3 we study here.

283 **tRF-3 targets.** miRNAs interact with their target mRNAs by base pairing. In plants, perfect
284 complementarity is essential for miRNA-mRNA targeting. In metazoans, with few exceptions, miRNA-
285 mRNA base pairing occurs imperfectly allowing unmatched bulges between the two RNAs. The
286 miRNA-mRNA base pairing rules are defined based on both bioinformatics and experimental results.
287 The most important rule is that miRNAs recognize their mRNA targets using their 2-7 (or 8) nt sequence
288 at their 5' ends, which is defined as the seed sequence. Second, an unpaired region at the middle of
289 miRNA-mRNA pairing (bulge) is tolerated. Third, following the bulge there is often sequence
290 complementarity between miRNA and mRNA ((Brennecke et al. 2005; Lewis et al. 2005; Grimson et al.
291 2007; Nielsen et al. 2007) and reviewed in (Filipowicz et al. 2008)). We show that tRF-3s down-regulate
292 target gene expression in a sequence dependent manner similar to miRNAs. Moreover, tRF-3:mRNA
293 pairs present in the Ago containing RISC also interact with GW proteins, mainly TNRC6A. Importantly,
294 our qRT-PCR analysis on luciferase reporters and RNA-seq analysis on endogenous genes show that
295 tRF-3s repress mRNA levels.

296 **The roles of Drosha, Dicer and Argonaute proteins.** The analysis of publicly available short
297 RNA sequences from *Dicer* knock out mouse embryonic stem cells showed that Dicer is not required for
298 generation of tRF-3s (Kumar et al. 2014; Kumar et al. 2015). In contrast, there are papers reporting that

299 Dicer might be necessary for tRF generation (Cole et al. 2009; Maute et al. 2013). It has been suggested
300 that tRNA may fold in alternate structure and that the reported role of Dicer in tRF generation is due to
301 the alternate structure that makes the tRNA susceptible to Dicer (Schopman et al. 2010). We find,
302 however, that in multiple cell lines tRF-3 generation is Drosha- and Dicer-independent and yet
303 Argonaute proteins are essential for their function (Figure 5 and Supplementary Figure S6).
304 Remarkably, tRF-3s repress their targets better in *Dicer* knock out cells suggesting that Argonaute
305 proteins are more accessible for tRF-3 loading in the absence of miRNAs (Figure 5D). This result raises
306 the possibility that tRFs become even more important regulators of gene expression in the absence of
307 miRNAs. Downregulation of Dicer and/or Drosha have been correlated with a worse outcome in lung,
308 breast, skin, endometrial and ovarian cancer (reviewed in (Foulkes et al. 2014)). Our results suggest that
309 tRF-3s acquire more potency when there are lower levels of Dicer or Drosha so that we should
310 investigate whether tRFs are important for the poorer outcome in these tumors.

311 On the other hand, Dicer has an important role in loading microRNAs into Argonaute complexes
312 (Chendrimada et al. 2005; Gregory et al. 2005). The fact that tRF-3s can repress genes by a mechanism
313 dependent on Argonaute proteins in Dicer knock out cells suggests that Dicer is not essential for loading
314 of tRF-3s into Argonaute. It is still possible that the Dicer-independent loading of tRF-3s into Argonaute
315 is less efficient than the Dicer-dependent loading of microRNAs. However, our results open the
316 possibility that other short RNAs, if present at high concentration, may load on to Argonaute complexes
317 and repress gene expression. Similarly, transfected miRNAs can load into RISC and repress their targets
318 in the absence of Dicer in mammalian cells (Betancur and Tomari 2012). Hsp90 proteins help the
319 stability of unloaded Argonaute proteins and loading of miRNA to Argonaute in an ATP-dependent way
320 (reviewed in (Meister 2013)), and thus Hsp90 may be involved in the loading of tRFs on Argonaute
321 proteins by this Dicer-independent pathway. It is worth noting that agotrons, short introns interacting

322 with Argonaute proteins, were shown to be loaded into Argonaute independently from Dicer, suggesting
323 that such loading is not unique to tRFs (Hansen et al. 2016).

324 **Regulators of tRF-3 levels.** The lack of any requirement for Dicer or Drosha for tRF-3
325 production opens up the question of how tRF-3s are generated. Identification of the specific
326 enzyme/enzymes that are important for tRF generation will let us better understand tRF biology and
327 function. The tRF-3s that we studied here contain CCA sequence which indicates that they are generated
328 from mature tRNA. We hypothesize that any process or enzyme regulating tRNA synthesis, maturation
329 or charging, such as levels/activities of CCA addition enzyme, tRNA modifiers like tRNA
330 methlytransferases or aminoacyl-tRNA synthetases will affect the levels of tRF-3s. Indeed, multiple
331 tRNAs have been reported to be upregulated in breast cancer (Pavon-Eternod et al. 2009). Recently,
332 Goodarzi et al. identified two upregulated tRNAs which result in overexpression of pro-metastatic
333 proteins in metastatic breast cancer cells. Whether tRFs derived from these tRNAs are upregulated in
334 parallel to control gene expression and contribute to the phenotype should be investigated. Moreover,
335 the Myc oncogene was shown to upregulate PolIII transcripts including tRNAs (Gomez-Roman et al.
336 2003; Arabi et al. 2005; Grandori et al. 2005; Lin et al. 2012). Since overexpression of tRNA in this
337 paper resulted in over production of tRF-3s, it is possible that when tRNA expression is induced by c-
338 Myc, there will be more tRF-3s. In parallel, many other oncogenes are known to induce tRNAs.
339 Examples include Ras, Raf, EGF receptor and oncogenes like E6 and E7 from human papilloma virus
340 (that inhibit p53 and Rb, both known to repress tRNA transcription) (Grewal 2015). Thus, it is
341 interesting to speculate that these oncogenes may indirectly increase the levels of specific tRFs. In line
342 with this, there is growing literature that levels of tRNA derived fragments are significantly altered in
343 many cancers (Zheng et al. 2016). More work is clearly needed to establish (a) whether tRF levels are

344 systematically altered in cancers and (b) whether these small tRFs can regulate gene expression and the
345 phenotypes of the cancers.

346 As more tRFs are implicated in different functions, it is exciting to note that at least a sub-
347 fraction of the tRFs also regulate gene expression by utilizing Argonaute-GW182 mediated pathways
348 that were first discovered in the context of microRNAs.

349

350 **Material and Methods:**

351 **Cloning:**

352 tRNA genes including their upstream and downstream elements which presumably regulate their
353 expression levels were amplified with specific primers and cloned by infusion cloning into pcDNA3
354 vector (See Supplementary Table 1 for primers and Supplementary Table 2 for list of plasmids). The
355 genomic locations that are used for the expression of tRNAs are as follows: tRNA LeuAAG
356 chr16.tRNA16 chr16:22308161-22308710 (hg19), tRNA-CysGCA chr17.tRNA26: chr17:37310553-
357 37311015(hg19), and tRNA LeuTAA chr6.tRNA83 chr6:144537474-144537897 (hg19).

358 Synthetic oligonucleotides containing perfect complementarity to tRF-3s were cloned at the
359 3'UTR of Renilla luciferase gene in psi-CHECK2 (Promega) reporter plasmid by infusion cloning
360 between PmeI and XhoI sites for luciferase assays. 3'UTRs of endogenous genes were similarly cloned
361 into psi-CHECK2 (Supplementary Table 2).

362 **qRT-PCR analysis of tRNAs/tRF-3s:**

363 4 µg of tRNA overexpression plasmids were transfected into HEK293T cells in 6 cm plates
364 using Lipofectamine 2000 (Thermofisher). The cells were collected after 2 days of transfection and total
365 RNA was purified for further analysis.

366 For detection of tRNAs by qRT-PCR, total RNAs were purified with TRIzol (Ambion)
367 extraction. 1µg of total RNA was subjected to DNase treatment using RQ1 RNase free DNase from
368 Promega. cDNAs were generated using Super script III reverse transcriptase (Thermo Fisher Scientific)
369 using random hexamers as primers according to manufacturer's instructions. Quantitative PCR were
370 performed using Sybr green mixes and specific primers against tRNA. Levels were normalized to
371 U6snRNA gene. Please see Supplementary Table 1 for list of primers.

372 For detection of tRFs, 100 µg of total RNA was subjected to small RNA enrichment using
373 miRVANA miRNA purification kit (Ambion). Small RNA enriched RNA pool was loaded into 15% 7M
374 Urea-Polyacrylamide page and RNAs in the 15-35 nt size range were purified. The size selected RNA
375 was eluted from gel slices overnight in 0.3 M NaCl/TE buffer at room temperature and precipitated with
376 2 volumes of isopropanol + 10 µg glycogen at -20°C overnight. RNA was precipitated by centrifugation
377 at 4°C for 20 min at 13,000 rpm and washed once with 70% EtOH. cDNAs were generated using
378 NCode miRNA First-Strand cDNA Synthesis and qRT-PCR Kits with 50-100ng 15-35 nt long RNA as
379 starting material. The levels of tRFs are normalized to the levels of miR-21. Please see Supplementary
380 Table 1 for the list of primers.

381 **Northern blotting:**

382 Cells were transfected and collected as described above.

383 40 µg of TRIzol extracted total RNA was loaded into 15% 7 M Urea-polyacrylamide gels. The
384 RNA was transferred onto Hybond-N+ positively charged nylon membranes (GE healthcare, USA Cat
385 No:RPN203B) using a Bio-Rad transblot apparatus at 200 mA (9 -10 V) for 3 hours (Bio-Rad, USA).
386 The transferred RNA was cross-linked to the membrane by UV-irradiation for 1 min (Stratalinker,
387 Stratagene) with 254-nm bulbs with autocrosslink option). The membrane was dried between two dry
388 3M papers by baking at 50°C for 30 min. The membrane was stored at 4°C between filter papers until

389 use. The rest of the protocol was performed as described in (Huang et al. 2014). Briefly, membranes
390 were pre-hybridized for at least 30 minutes at 40°C in pre-hybridization buffer (7% SDS, 200 mM
391 Na2HPO4 (pH 7.0)) containing 5 µg/ml denatured salmon sperm DNA. Hybridization was performed in
392 Expresshyb solution, Clontech Cat No: 636831 containing 50 pmol/ml labeled probes (Anti-3009: BIO-
393 TGGTACCAGGAGTGGGGT) at 40°C overnight (typically 16 hrs). The membrane was washed thrice
394 with 1X SSC, 0.1 % SDS at RT for 5 min. ECL Lightning was performed using Chemiluminescent
395 Nucleic Acid Detection Module Kit from Thermo Scientific, USA following the instructions in their
396 manual.

397 **Dual Luciferase assay:**

398 Luciferase assays were performed in 24-well plates. 480 ng of tRNA overexpression plasmid or
399 empty vector pcDNA3 and 2 ng of psi-CHECK2 reporter vector with no sites or perfect complementary
400 sites to tRF-3 in the 3'UTR of Renilla luciferase gene were transfected in to HEK293T cells using
401 Lipofectamine 2000 (Life Technologies) as a transfection reagent. The cells were lysed in 100 ul 1X
402 passive lysis buffer from Dual-Luciferase® Reporter Assay System (Promega) after 2 days of
403 transfection. The luciferase signals were measured using 20 ul of the lysate following the instructions
404 provided by Promega. The Renilla luciferase levels were normalized to firefly luciferase levels and the
405 results were always plotted with tRF overexpression over non-overexpression (see figure legends).

406 Luciferase reporter assays were done in the same way with reporters containing 3'UTRs of
407 endogenous genes.

408

409 **siRNA knock down experiments:**

410 To knock down Argonaute proteins, 20 nM of siControl(Sigma) or 20 nM total of siRNAs
411 against Ago1, 2 and 3 were transfected into HEK293T cells using RNAimax transfection reagent from

412 Life Technologies, and this was repeated after 24hrs. Luciferase reporters and tRNA overexpression
413 plasmids were transfected 24 hrs after second siRNA transfection. Lysates were collected 48 hr after
414 second siRNA transfection for western blot analysis and dual luciferase assay.

415 **Western blotting:**

416 Argonaute protein levels were detected by Western blotting. Briefly, the membrane was
417 incubated in 5% milk in TBS-T blocking solution for 1 hr at room temperature and in rabbit monoclonal
418 primary (Ago1: CST#5053, Ago2: CST#2897) in 5% BSA, TBS-T with 1:1000 dilution for overnight at
419 4°C. The membrane was washed 3 times in 1X TBS-T and incubated with 1:5000 diluted HRP goat-
420 anti-rabbit secondary in 5% BSA, TBST at room temperature. Immobilon Western Chemiluminescent
421 HRP substrate from Millipore was used for developing the signal.

422 20ug of cell lysate from wild type and Dicer knockout cells was loaded into 7.5% SDS-PAGE
423 gel to detected Dicer protein levels by western blotting. Blocking was performed in 3% milk in PBS-T
424 for 1hr at room temperature. 1:1000 dilution of Dicer antibody (Abcam ab14601) and 1:2000 dilution of
425 α-tubulin antibody (Santa Cruz sc-5286) was used as a primary at 4°C for overnight.

426 **AGO Immunoprecipitation:**

427 Immunoprecipitation of Argonaute proteins from tRNA over-expressing cells was performed as
428 described in MAGNA-RIP kit (Millipore 17-700) using Pan-Ago antibody (Millipore MABE56). RNA
429 was extracted by phenol chloroform as described in the kit and loaded into 15% 7 M Urea-
430 polyacrylamide gel. Northern was performed as described above.

431 **RNA-seq library preparation:**

432 HEK293T cells were transfected with empty vector or tRNA expression plasmids 4 times in two
433 days intervals. Total RNA was purified with Qiagen RNeasy kit. tRNA and tRF levels were quantified
434 as described above.

435 1 ug of total RNA from pcDNA3 and tRNA LeuTAA (tRF-3009) overexpressing cells were used
436 for library preparation. Libraries were prepared using NEBNext Ultra directional RNA library prep kit
437 with NEBNext Poly(A) mRNA Magnetic Isolation Module for Illumina. The libraries were indexed
438 using NEBNext Multiplex oligos for Illumina. Quality controls and sequencing was done in Genomic
439 Services Lab at Hudson Alpha.

440 **Analysis of the small RNA data isolated from Dicer knock out and Dicer WT cells:**

441 We analyzed two small RNA datasets isolated from Dicer knock out and wild type cell lines
442 generated by two independent laboratories (Bogerd et al. 2014; Kim et al. 2016). Firstly, adaptor
443 sequence was removed using the `Cutadapt' program (Martin 2011) and sequencing reads that were
444 >=14 bases long were retained. To identify the total mapped reads the small RNA reads were mapped on
445 whole genome (hg38 genome build) by using short read aligner Novoalign
446 (http://www.novocraft.com/products/novoalign/). Next, the identical reads were collapsed and mapped
447 on to the in house built small RNA database (mature miRNA and tRNA as detailed in (Kumar et al.
448 2014; Kumar et al. 2015)) using BLASTn (Altschul et al. 1990). The building of in house blast database
449 for blast searches is explained in detail in our earlier publications (Kumar et al. 2014; Kumar et al.
450 2015). The total number of mapped reads was used for normalizing the expression of miRNA and tRFs.
451 In general we considered only those alignments where the query sequence (small RNA) was mapped to
452 the database sequence (tRNA or miRNA) along 100% of its length. The blast output file was parsed to
453 get information on the mapped position of small RNA on tRNA or miRNA. We extracted all map
454 positions where the small RNA aligned from its first base to the last base with the tRNA sequence
455 allowing either one or no mismatch. Since 'CCA' is added at the 3' end of tRNA by tRNA
456 nucleotidyltransferase during maturation of tRNA (Xiong and Steitz 2006), we allowed a special
457 exception for the small RNA mapping to the 3' ends of tRNAs allowing a terminal mismatch of < =3

458 bases. To remove any false positives, the small RNAs that mapped on to the 'tRNAdb' were again
459 searched against the whole genome using blast search excluding the tRNA loci.

460 **Analysis of PAR CLIP data:**

461 We also investigated tRF and miRNA expression in human small RNA PARCLIP data of
462 TNRC6A, B & C by analyzing the human TNRC6A (GEO ID = GSM545218), B (GEO ID =
463 GSM545219), and C (GEO ID = GSM545220) PAR-CLIP data isolated from HEK293 cell lines
464 (Hafner et al. 2010). Data from all three small RNA libraries were examined for miRNA and tRF
465 expression as well as for the T to C mutation position and its frequency compared to wild type small
466 RNAs (miRNAs and tRFs). Sequence reads that either mapped perfectly on miRNA or tRFs or mapped
467 with one base mismatch were considered for T to C mutation analysis. Mismatched base and its position
468 relative to the 5' end of small RNA were collected for final analysis.

469 The 17,319 crosslink-centered regions (CCRs) identified by Hafner et al (Hafner et al. 2010)
470 present in the PAR-CLIP data was used to study the complementary sequence of miRNA and tRF-3 seed
471 sequence along the length of CCR. All the possible 7-mer sequence was generated along the length of
472 miRNA and tRF. The 7-mer sequences were reverse complemented and mapped and the match was
473 scored along the length of CCR. CCRs are 41 nt long sequences centered at the T (protein binding site)
474 that showed the highest T to C frequency. Hafner et al. demonstrated that the reverse complement of
475 known miRNA seeds is enriched in CCRs directly following this central cross-linked T. In our analysis
476 four 7-mer (1-7, 2-8, 3-9 & 4-10) reverse complementary sequence of top 20 abundant miRNA and tRF-
477 3 sequences identified in TNRC6A-C (a member of GW-bodies or P-bodies) were used for finding
478 sequences along the length of CCR. The counting and scanning of the CCRs was done from the 5' end

479 to 3' end of CCRs. Whenever there was a match, the score (count) was assigned to all the seven bases of
480 CCR.

481 **Analysis RNASeq data:**

482 We received on an average 30 million 50 bases long paired end reads for each of the replicates
483 and had two replicates for each condition. The transcript (RefSeq genes) sequences for the genome
484 build hg38 were downloaded from UCSC table browser on Dec 10, 2016 (http://genome.ucsc.edu). We
485 used default parameters of Kallisto (Bray et al. 2016) to build an index for the above transcript sequence
486 and then quantified abundances of the transcripts from the paired end RNAseq fastq reads (Bray et al.
487 2016). The DESeq2 (Love et al. 2014) package in R was used for differential expression analysis of the
488 quantified data obtained from Kallisto. The normalized count data from DESeq2 was used for all other
489 downstream analysis. The data has been deposited to Gene Expression Omnibus (GEO)
490 database,www.ncbi.nlm.nih.gov/geo.

491 **tRF target gene prediction:**

492 The 3'UTR sequence of each annotated RefSeq genes was downloaded using UCSC Table
493 Browser [hg38 genome build]. We decrease the noise from the lowly expressed isoform by considering
494 only most expressed isoform of each genes in Hela cells (Nam et al. 2014) for downstream analysis. A
495 total of 9294 sequences were examined for the complementarity of tRF-3009 sequence using the default
496 parameter RNA22 (Miranda et al. 2006). 1119 3'UTR sequences were identified that had at least 6-mer
497 complementary sequence to 5' region of the tRF-3009 (Miranda et al. 2006). These identified targets
498 were used to compare the expression of target with non-target genes.

499 **Cumulative distribution function plot (CDF plot) of tRF target and non-target genes**

500 Cumulative distribution function of R (ecdf) was used to compare the plot between targets and
501 non-target genes in various experimental conditions. ks.test (Kolmogrov-Smirnov test), a function in R
502 package was used to test if the plot of target genes is above the plot of non-target genes and the
503 difference is statistically significant.

504 **Author contributions:** C.K. and A.D. designed the experiments. C.K. is responsible for all biological
505 experiments and P.K. is responsible for bioinformatics analysis. Initial bioinformatics analysis was done
506 by M.K. A.M performed Ago western blots. Z.S. performed some of the biological experiments. C.K.
507 and A.D. wrote the paper.

508 **Acknowledgements:**

509 We would like to thank all the members of Dutta lab for fruitful discussions. We also would like
510 to thank Dr. Bryan Cullen for the Dicer knock out 293T cells. This work was supported by P01
511 CA104106 (Project 3) and R01 GM84465 to A.D. M.K. was partly supported by a DOD award
512 (PC151085).

513 **References:**

514 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol*
515 *Biol* **215**: 403-410.
516 Arabi A, Wu S, Ridderstrale K, Bierhoff H, Shiue C, Fatyol K, Fahlen S, Hydbring P, Soderberg O,
517 Grummt I et al. 2005. c-Myc associates with ribosomal DNA and activates RNA polymerase I
518 transcription. *Nature cell biology* **7**: 303-310.
519 Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the
520 initiation step of RNA interference. *Nature* **409**: 363-366.
521 Betancur JG, Tomari Y. 2012. Dicer is dispensable for asymmetric RISC loading in mammals. *Rna-a*
522 *Publication of the Rna Society* **18**: 24-30.
523 Bogerd HP, Whisnant AW, Kennedy EM, Flores O, Cullen BR. 2014. Derivation and characterization of
524 Dicer- and microRNA-deficient human cells. *RNA* **20**: 923-937.
525 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification.
526 *Nat Biotechnol* **34**: 525-527.
527 Brennecke J, Stark A, Russell RB, Cohen SM. 2005. Principles of microRNA-target recognition. *PLoS*
528 *biology* **3**: e85.
529 Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. 2010. A dicer-independent miRNA biogenesis
530 pathway that requires Ago catalysis. *Nature* **465**: 584-589.
531 Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, Feng GH, Peng H, Zhang X, Qian J et al. 2016. Sperm
532 tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science*
533 **351**: 397-400.
534 Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, Shiekhattar R. 2005.
535 TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*
536 **436**: 740-744.
537 Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, Brown JW, Green PJ, Barton GJ, Hutvagner G.
538 2009. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small
539 RNAs derived from tRNAs. *RNA* **15**: 2147-2160.
540 Cozen AE, Quartley E, Holmes AD, Hrabeta-Robinson E, Phizicky EM, Lowe TM. 2015. ARM-seq:
541 AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA
542 fragments. *Nature methods* **12**: 879-884.
543 Eulalio A, Huntzinger E, Izaurralde E. 2008. GW182 interaction with Argonaute is essential for
544 miRNA-mediated translational repression and mRNA decay. *Nature structural & molecular*
545 *biology* **15**: 346-353.
546 Filipowicz W, Bhattacharyya SN, Sonenberg N. 2008. Mechanisms of post-transcriptional regulation by
547 microRNAs: are the answers in sight? *Nature reviews Genetics* **9**: 102-114.
548 Foulkes WD, Priest JR, Duchaine TF. 2014. DICER1: mutations, microRNAs and mechanisms. *Nature*
549 *reviews Cancer* **14**: 662-672.
550 Gomez-Roman N, Grandori C, Eisenman RN, White RJ. 2003. Direct activation of RNA polymerase III
551 transcription by c-Myc. *Nature* **421**: 290-294.
552 Grandori C, Gomez-Roman N, Felton-Edkins ZA, Ngouenet C, Galloway DA, Eisenman RN, White RJ.
553 2005. c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by
554 RNA polymerase I. *Nature cell biology* **7**: 311-318.
555 Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. 2005. Human RISC couples microRNA
556 biogenesis and posttranscriptional gene silencing. *Cell* **123**: 631-640.

557 Grewal SS. 2015. Why should cancer biologists care about tRNAs? tRNA synthesis, mRNA translation
558 and the control of growth. *Biochimica et biophysica acta* **1849**: 898-907.
559 Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting
560 specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**: 91-105.
561 Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr.,
562 Jungkamp AC, Munschauer M et al. 2010. Transcriptome-wide identification of RNA-binding
563 protein and microRNA target sites by PAR-CLIP. *Cell* **141**: 129-141.
564 Hansen TB, Veno MT, Jensen TI, Schaefer A, Damgaard CK, Kjems J. 2016. Argonaute-associated
565 short introns are a novel class of gene regulators. *Nature communications* **7**: 11538.
566 Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA. 2010. Human tRNA-derived small
567 RNAs in the global regulation of RNA silencing. *RNA* **16**: 673-695.
568 Helwak A, Kudla G, Dudnakova T, Tollervey D. 2013. Mapping the human miRNA interactome by
569 CLASH reveals frequent noncanonical binding. *Cell* **153**: 654-665.
570 Honda S, Morichika K, Kirino Y. 2016. Selective amplification and sequencing of cyclic phosphate-
571 containing RNAs by the cP-RNA-seq method. *Nature protocols* **11**: 476-489.
572 Huang Q, Mao Z, Li S, Hu J, Zhu Y. 2014. A non-radioactive method for small RNA detection by
573 northern blotting. *Rice (N Y)* **7**: 26.
574 Jonas S, Izaurralde E. 2015. Towards a molecular understanding of microRNA-mediated gene silencing.
575 *Nature reviews Genetics* **16**: 421-433.
576 Keam SP, Hutvagner G. 2015. tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient
577 RNA in the Regulation of Gene Expression. *Life (Basel)* **5**: 1638-1651.
578 Kim HK, Fuchs G, Wang S, Wei W, Zhang Y, Park H, Roy-Chaudhuri B, Li P, Xu J, Chu K et al. 2017.
579 A transfer-RNA-derived small RNA regulates ribosome biogenesis. *Nature* **552**: 57-62.
580 Kim YK, Kim B, Kim VN. 2016. Re-evaluation of the roles of DROSHA, Export in 5, and DICER in
581 microRNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of*
582 *America* **113**: E1881-1889.
583 Kumar P, Anaya J, Mudunuri SB, Dutta A. 2014. Meta-analysis of tRNA derived RNA fragments
584 reveals that they are evolutionarily conserved and associate with AGO proteins to recognize
585 specific RNA targets. *BMC biology* **12**: 78.
586 Kumar P, Kuscu C, Dutta A. 2016. Biogenesis and Function of Transfer RNA-Related Fragments
587 (tRFs). *Trends in biochemical sciences* **41**: 679-689.
588 Kumar P, Mudunuri SB, Anaya J, Dutta A. 2015. tRFdb: a database for transfer RNA fragments.
589 *Nucleic acids research* **43**: D141-145.
590 Lee YS, Shibata Y, Malhotra A, Dutta A. 2009. A novel class of small RNAs: tRNA-derived RNA
591 fragments (tRFs). *Genes & development* **23**: 2639-2649.
592 Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates
593 that thousands of human genes are microRNA targets. *Cell* **120**: 15-20.
594 Lin CY, Loven J, Rahl PB, Paranal RM, Burge CB, Bradner JE, Lee TI, Young RA. 2012.
595 Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**: 56-67.
596 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq
597 data with DESeq2. *Genome Biol* **15**: 550.
598 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
599 *EMBnetjournal, North America* **17**: 3.
600 Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, Dalla-Favera R. 2013. tRNA-
601 derived microRNA modulates proliferation and the DNA damage response and is down-

regulated in B cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 1404-1409.

Meister G. 2013. Argonaute proteins: functional insights and emerging roles. *Nature reviews Genetics* **14**: 447-459.

Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I. 2006. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* **126**: 1203-1217.

Nam JW, Rissland OS, Koppstein D, Abreu-Goodger C, Jan CH, Agarwal V, Yildirim MA, Rodriguez A, Bartel DP. 2014. Global analyses of the effect of different cellular contexts on microRNA targeting. *Molecular cell* **53**: 1031-1043.

Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB. 2007. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* **13**: 1894-1910.

Pavon-Eternod M, Gomes S, Geslain R, Dai Q, Rosner MR, Pan T. 2009. tRNA over-expression in breast cancer and functional consequences. *Nucleic acids research* **37**: 7268-7280.

Pliatsika V, Loher P, Magee R, Telonis AG, Londin E, Shigematsu M, Kirino Y, Rigoutsos I. 2018. MINTbase v2.0: a comprehensive database for tRNA-derived fragments that includes nuclear and mitochondrial fragments from all The Cancer Genome Atlas projects. *Nucleic acids research* **46**: D152-D159.

Saikia M, Hatzoglou M. 2015. The Many Virtues of tRNA-derived Stress-induced RNAs (tiRNAs): Discovering Novel Mechanisms of Stress Response and Effect on Human Health. *The Journal of biological chemistry* **290**: 29761-29768.

Schopman NC, Heynen S, Haasnoot J, Berkhout B. 2010. A miRNA-tRNA mix-up: tRNA origin of proposed miRNA. *RNA biology* **7**: 573-576.

Schorn AJ, Gutbrod MJ, LeBlanc C, Martienssen R. 2017. LTR-Retrotransposon Control by tRNA-Derived Small RNAs. *Cell* **170**: 61-71 e11.

Selitsky SR, Sethupathy P. 2015. tDRmapper: challenges and solutions to mapping, naming, and quantifying tRNA-derived RNAs from human small RNA-sequencing data. *BMC bioinformatics* **16**: 354.

Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Belleannee C, Kucukural A, Serra RW, Sun F et al. 2016. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**: 391-396.

Telonis AG, Loher P, Honda S, Jing Y, Palazzo J, Kirino Y, Rigoutsos I. 2015. Dissecting tRNA-derived fragment complexities using personalized transcriptomes reveals novel fragment classes and unexpected dependencies. *Oncotarget* **6**: 24797-24822.

Thompson A, Zielezinski A, Plewka P, Szymanski M, Nuc P, Szweykowska-Kulinska Z, Jarmolowski A, Karlowski WM. 2018. tRex: A Web Portal for Exploration of tRNA-Derived Fragments in Arabidopsis thaliana. *Plant & cell physiology* **59**: e1.

Xiong Y, Steitz TA. 2006. A story with a good ending: tRNA 3'-end maturation by CCA-adding enzymes. *Curr Opin Struct Biol* **16**: 12-17.

Zheng G, Qin Y, Clark WC, Dai Q, Yi C, He C, Lambowitz AM, Pan T. 2015. Efficient and quantitative high-throughput tRNA sequencing. *Nature methods* **12**: 835-837.

Zheng LL, Xu WL, Liu S, Sun WJ, Li JH, Wu J, Yang JH, Qu LH. 2016. tRF2Cancer: A web server to detect tRNA-derived small RNA fragments (tRFs) and their expression in multiple cancers. *Nucleic acids research* **44**: W185-193.

647 **Figure Legends:**

648

649 **Figure 1: tRF-3s have distinct lengths and interact with Argonaute (A)** tRNA secondary structure
650 depicting the tRF-3 cleavage sites. **(B)** Read counts for tRF-3s in AGO PAR-CLIP data (Hafner *et al*.,
651 2010). Individual tRF-3s are arrayed along the X-axis with their expression levels (Number of reads per
652 million mapped reads) shown on the Y-axis. The tRF-3s studied in this paper are indicated. **(C)** Number
653 of reads per million mapped reads for each tRF-3 in top 300 chimeric reads from Ago-CLASH data
654 (Helwak et al., 2013) **(D)** Mapped position along the length of the parental tRNA (X-axis) of small
655 RNAs derived from that tRNA and their abundance (Y-axis: number of reads found in
656 library GSM416733). Small RNAs from tRNA LeuAAG chr16.tRNA16 (upper), tRNA-CysGCA
657 chr17.tRNA26 (middle), and tRNA LeuTAA chr6.tRNA83 (lower) are shown and the tRFs studied in
658 this paper indicated. X-axis is the position on tRNA gene. Blue arrowhead and *** indicate the end of
659 mature tRNA and anticodon, respectively.

660

661 **Figure 2: tRF-3s produced by tRNA overexpression are loaded into Argonaute. (A)** Relative levels
662 of indicated tRNAs upon tRNA overexpression. Mean and s.d. of three independent experiments. *: p-
663 value <0.05 (Wilcoxon-Mann-Whitney Test). **(B)** Relative levels of indicated tRFs upon overexpression
664 of the tRNAs indicated in (A) in the same order from left to right. Mean and s.d. of at least three
665 independent experiments. *: p-value <0.05 (Wilcoxon-Mann-Whitney Test). **(C)** Northern blot showing
666 the tRF-3s produced after overexpression of the indicated parental tRNA. Lower panel shows equal
667 loading of lanes. **(D)** Northern blot showing the association with Argonaute of tRF-3009a produced
668 from tRNA overexpression. The immunoblot showing successful Argonaute immunoprecipitation is in
669 Supplementary Figure S2.

670

671 **Figure 3: tRF-3s down regulate target expression through complementarity in 3'UTR of luciferase**
672 **reporter. (A)** Luciferase reporter assays using Renilla luciferase with a perfect complementary
673 sequence to tRF-3 at the 3'UTR. Renilla luciferase levels were first normalized to Firefly luciferase
674 levels from the same transfection and then normalized to no tRNA/tRF overexpression (empty vector)
675 control. *: p-value <0.05 (t-test). **(B)** Degree of repression correlates with tRNA overexpression amount.
676 Amount of transfected tRNA plasmid was titrated to measure the change in the degree of repression.
677 Luciferase reporter assays were analyzed as described in Fig. 3A. **(C)** Luciferase reporter assay showing
678 the specific repression by each corresponding tRF-3. (*: p-value <0.05 (t-test)).

679

680 **Figure 4: Seed sequence is required for target repression by tRF-3s.** Luciferase reporter assays with
681 mutant target site at the luciferase reporter upon tRF-3001 **(A)**, tRF-3003 **(B)** and tRF-3009 **(C)**
682 overexpression. Canonical seed region on each tRF-3 and complementary sequence on each target are
683 highlighted in yellow. Mutated regions are underlined and colored red. P values are calculated by t-test
684 comparing luciferase reporter with indicated target sequence to empty vector control (*: p-value < 0.05,
685 ** : p-value < 0.005 (t-test)).

686

687

688 **Figure 5: Target repression by tRF-3s is independent of Dicer but dependent on Argonautes. (A)**
689 Luciferase reporter assays after tRNA overexpression to produce tRF-3009 ± knockdown of Argonaute
690 proteins. *: p-value < 0.05 (t-test). **(B)** Northern blot showing tRF-3009 levels in tRNA overexpressing
691 cells after Ago knock down. Lower panel shows equal loading of lanes. **(C)** tRF-3 read counts in small
692 RNA sequencing data from WT and two different Dicer knockout clones of HEK293T. Small RNA

693 sequencing data from (Bogerd et al. 2014) **(D)** Luciferase reporter assays in wild type and Dicer knock
694 out HEK293T cells, NoDice 4-25 (middle) and NoDice 2-20 (right) (Bogerd et al. 2014).
695

696 **Figure 6: tRF-3s associate with GW182/TNRC6 proteins. (A)** Read counts for miRNAs or tRF-3s in
697 TNRC6A, TNRC6B and TNRC6C PAR-CLIP data from (Hafner et al. 2010). Each microRNA or tRF is
698 given an arbitrary identifying number along the X-axis. The expression level (expressed as reads per
699 million mapped reads) of each microRNA or tRF is shown on the Y-axis. **(B)** Normalized positional T to
700 C mutation frequencies for miRNA or tRF-3 reads found in the TNRC6A, TNRC6B and TNRC6C
701 PAR-CLIP data. **(C-D)** Complementarity in the target RNA CCRs present in the TNRC6A PAR-CLIP
702 to the 1-7, 2-8 3-9 and 4-10-mer sequences from 5' end of 20 most abundant microRNAs or tRF-3s seen
703 in (B). The CCRs are centered on the site of the U/C mutation and the number of targets with
704 complementarity indicated at the corresponding base in the sequence of the CCR. For example, the
705 maximum number of matches is seen with the 1-7 mer of the tRF and begins with the base immediately
706 downstream from the cross-link site in the target RNA.
707

708 **Figure 7: Endogenous targets are repressed upon tRF-3009 overexpresssion by tRNA-LeuTAA**
709 **transfection. (A)** Luciferase assay upon tRNA overexpression to produce tRF-3009 and Renilla
710 luciferase mRNA levels detected by qRT-PCR from the same cells that luciferase assay performed
711 (normalized to Firefly mRNA levels) (*: p-value < 0.005) **(B)** Cumulative distribution function (CDF)
712 plots showing the repression of tRF-3009 targets upon overexpression of tRNA producing tRF-3009.
713 Targets have been predicted using RNA22 algorithm (Miranda et al, 2006). **(C)** Seed sequence
714 complementarity in selected tRF-3009 targets that are identified in RNA-seq upon tRF-3009 expression.
715 The number after 3'UTR indicates the start position of the mRNA sequence match with 1 being the base

716 immediately downstream from the stop codon. Red line: perfect base-pairing in seed; black line: perfect
717 base-pairing outside seed; dashed line: wobble base-pairing. FER1 gene has two predicted
718 complementary sites on its 3' UTR. (D) Relative mRNA levels of indicated genes upon tRNA-LeuTAA
719 transfection, leading to tRF-3009 overexpression (*: p-value < 0.05, **: p-value, 0.005 (t-test)). **(E)**
720 Dual luciferase reporter assay on reporters containing the 3'UTR of indicated genes after tRNA-
721 LeuTAA transfection, leading to tRF-3009 overexpression. Perfect complementary sequence to the tRF-
722 3009 serves as a positive control and all results are normalized to the "no site" reporter without any
723 match to the tRF-3009 (*: p-value < 0.05, **: p-value, 0.005 (t-test)).

724 **Supplementary Figure S1:** Sequencing coverage (normalized to reads per millions reads in library) of
725 each of the bases of tRNA based on reads mapped on to tRNA LeuAAG chr16.tRNA16 (upper), tRNA-
726 CysGCA chr17.tRNA26 (middle), and tRNA LeuTAA chr6.tRNA83 (lower). X-axis is the position on
727 tRNA gene. Y axis is the number of times a particular base has been sequenced in GSM416733. Blue
728 arrowhead and *** indicate the end of mature tRNA and anticodon, respectively.

729 **Supplementary Figure S2: Immunoprecipitation of Argonaute.** Western blotting showing the
730 specific pull down of Argonaute in the RIP experiment.

731

732 **Supplementary Figure S3: Degree of repression is correlated with amount of tRNA expression**
733 **plasmid.** Luciferase reporter assay after tRF-3003 **(A)** or tRF-3009 **(B)** overexpression. Amount of
734 transfected tRNA plasmid was titrated to show the change in the degree of repression. Renilla firefly
735 with a perfect complementary to tRF-3 sequence at the 3' UTR is used as a reporter. Renilla luciferase
736 levels normalized to Firefly luciferase levels and then normalized back to no tRNA/tRF overexpression
737 condition.

738

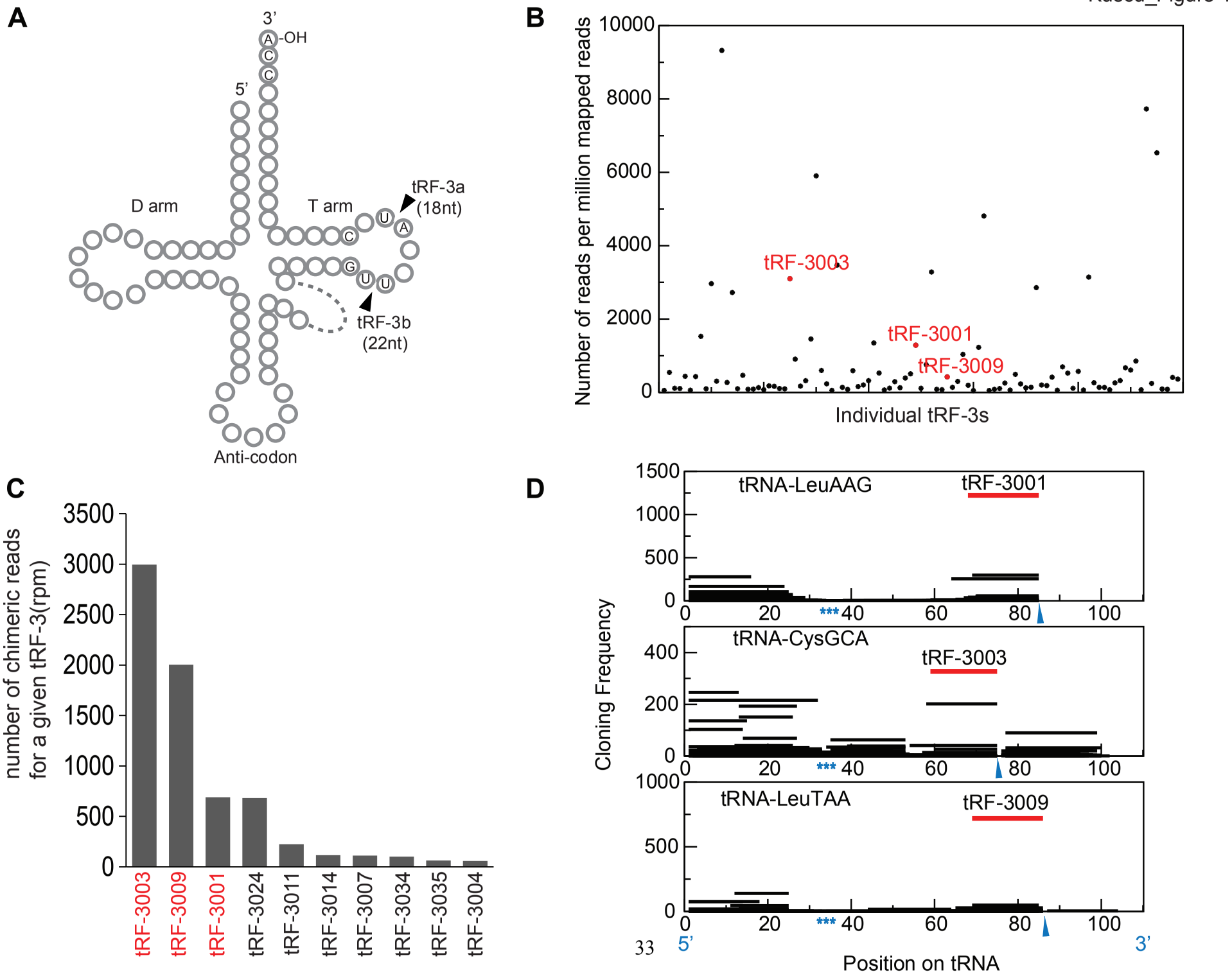
739 **Supplementary Figure S4: tRF-3009 mimic also showed repression by following similar rules to**
740 **tRNA-LeuTAA overexpression producing tRF-3009.** Luciferase reporter assays with target site on the
741 luciferase reporter after transfection of tRF-3009 mimic. The tRF-3009a mimic sequence is shown in
742 bold with the "Perfect complementary" target site below. Mutated regions in the various target plasmids
743 are underlined and colored red. P values are calculated by t-test comparing luciferase reporter with
744 indicated target sequence to empty vector control (*: p-value < 0.05, ** : p-value < 0.005 (t-test)).
745

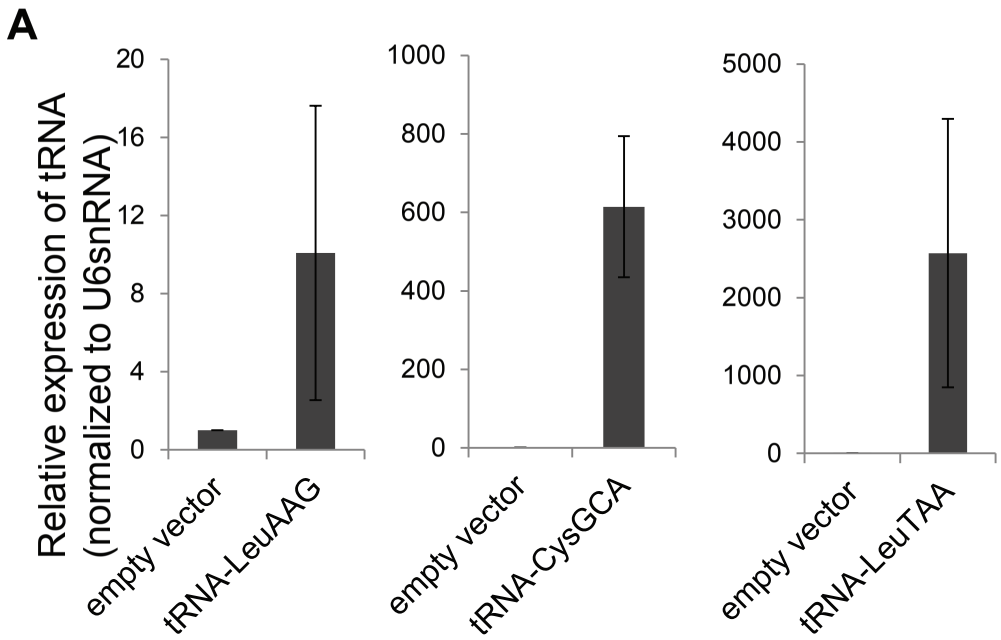
746 **Supplementary Figure S5: Ago is necessary for tRF-3 function. (A)** Western blot showing Argonaute
747 protein levels after siRNA knockdown. Luciferase reporter assays under Argounate proteins knockdown
748 upon tRF-3001 **(B)** or tRF-3003 **(C)** overexpression. *: p-value < 0.05 (t-test).

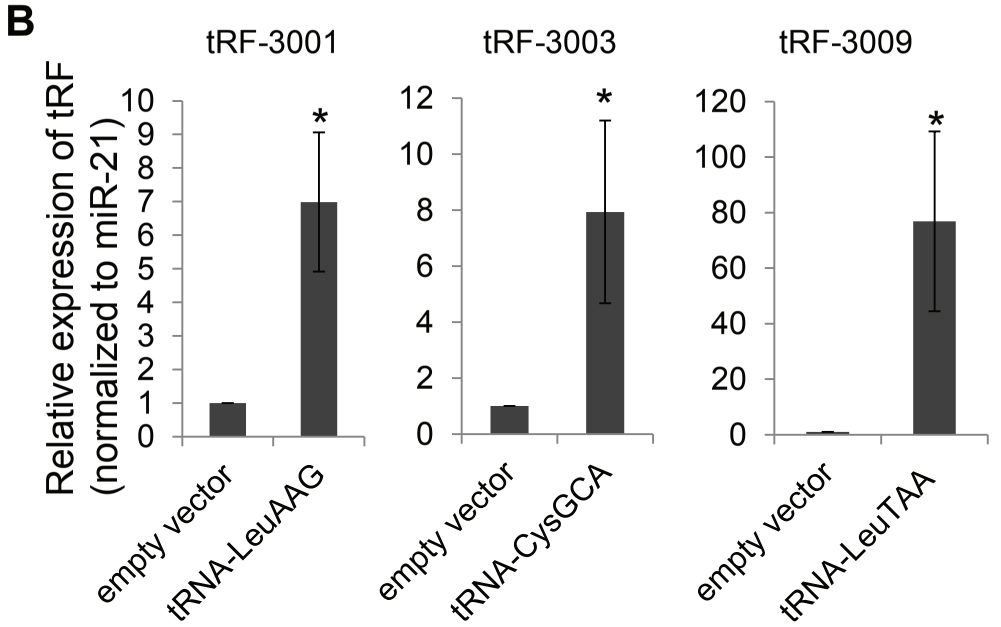
749 **Supplementary Figure S6: Role of Dicer, Drosha and Exportin 5 in tRF-3 generation (A)** Western
750 blots showing the levels of Dicer levels in WT and Dicer KO cells (Bogerd et al. 2014). **(B)** tRF-3 read
751 counts in small RNA sequencing data from WT and two different Dicer knockout clones of HEK293T.
752 Small RNA sequencing data from (Bogerd et al. 2014) miRNA **(C)** and tRF-3 **(D)** read counts in small
753 RNA sequencing data from WT, Drosha, Dicer and Exportin-5 knockout HCT116 cells (Kim et al.
754 2016).
755

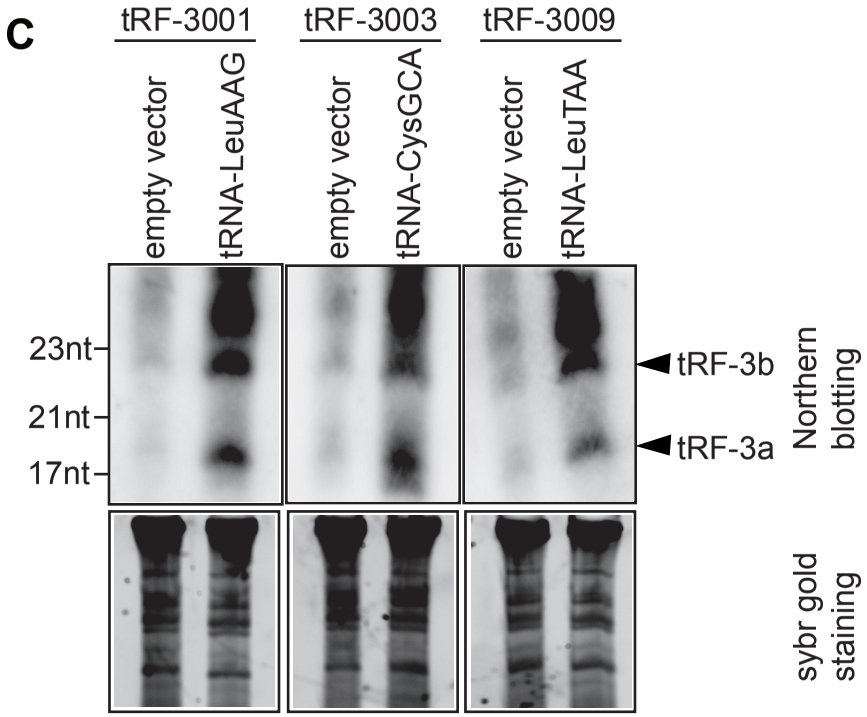
756 **Supplementary Figure S7: tRF-3009 represses its endogenous targets.** Cumulative distribution
757 function (CDF) plots showing the repression of tRF-3009 targets upon tRF-3009 overexpression from
758 the second biological replicate of the experiment.
759

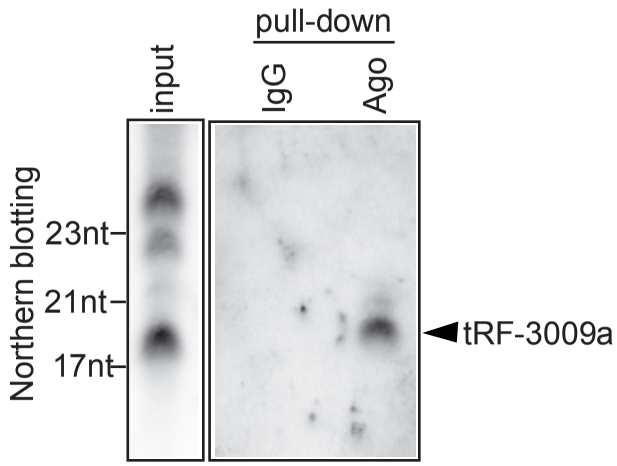
Kuscu_Figure 1



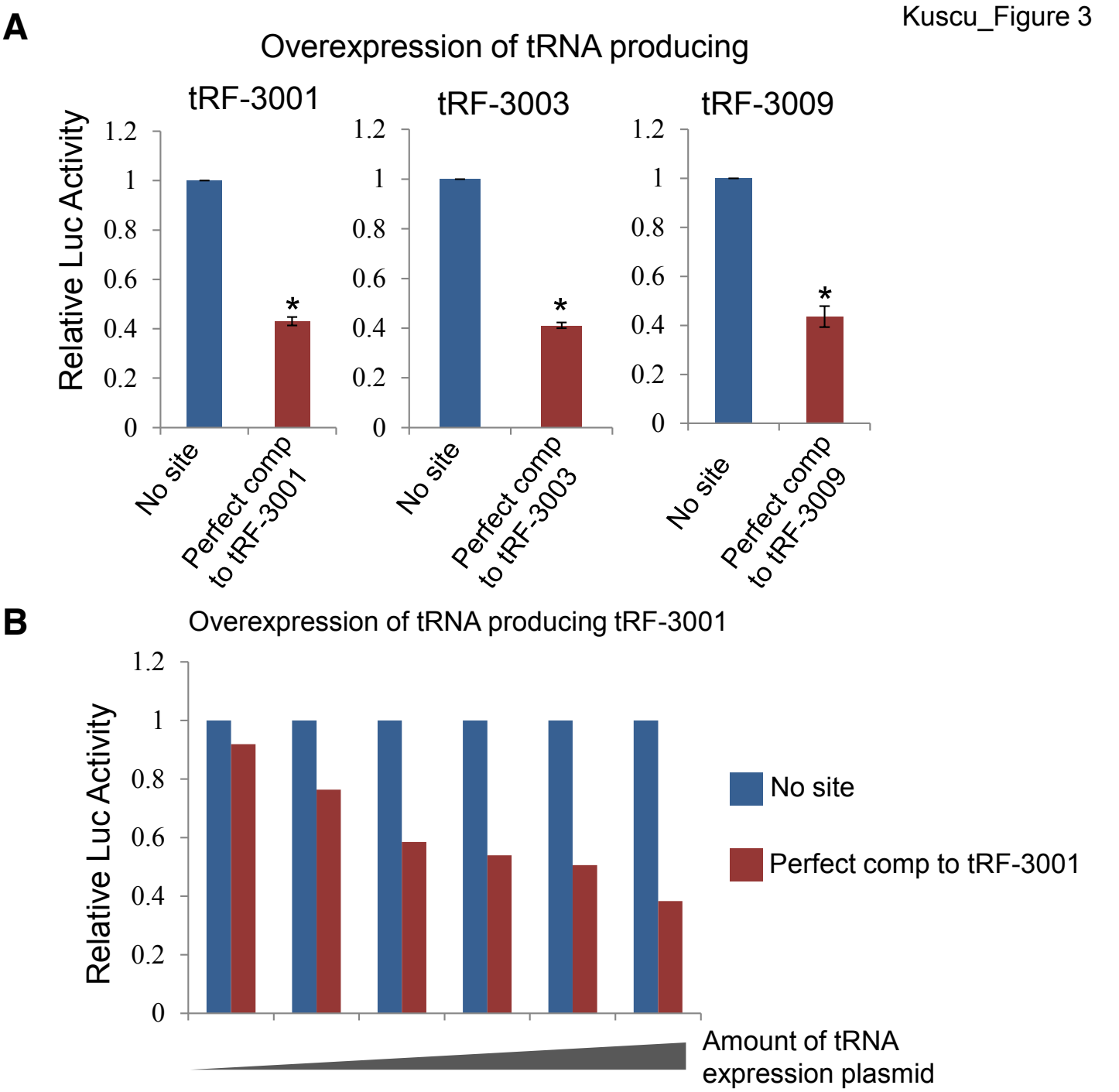

A
Relative expression of tRNA (normalized to U6snRNA)
20
16
12
8
4
0
1000
800
600
400
200
0
5000
4000
3000
2000
1000
0
empty vector
tRNA-LeuAAG
empty vector
tRNA-CysGCA
empty vector
tRNA-LeuTAA

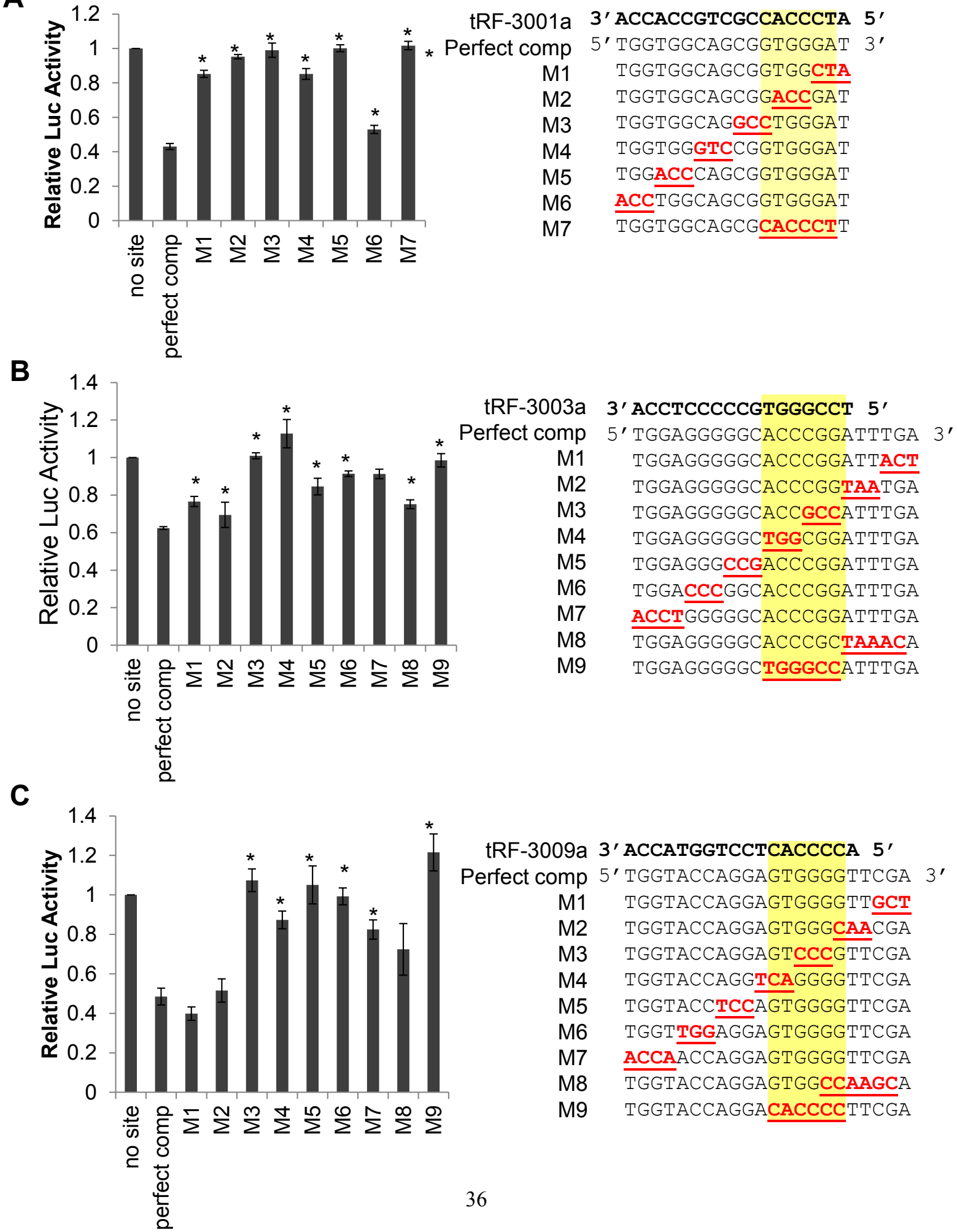

B
tRF-3001
tRF-3003
tRF-3009
Relative expression of tRF (normalized to miR-21)
10
9
8
7
6
5
4
3
2
1
0
12
10
8
6
4
2
0
120
100
80
60
40
20
0
*
*
*
empty vector
tRNA-LeuAAG
empty vector
tRNA-CysGCA
empty vector
tRNA-LeuTAA

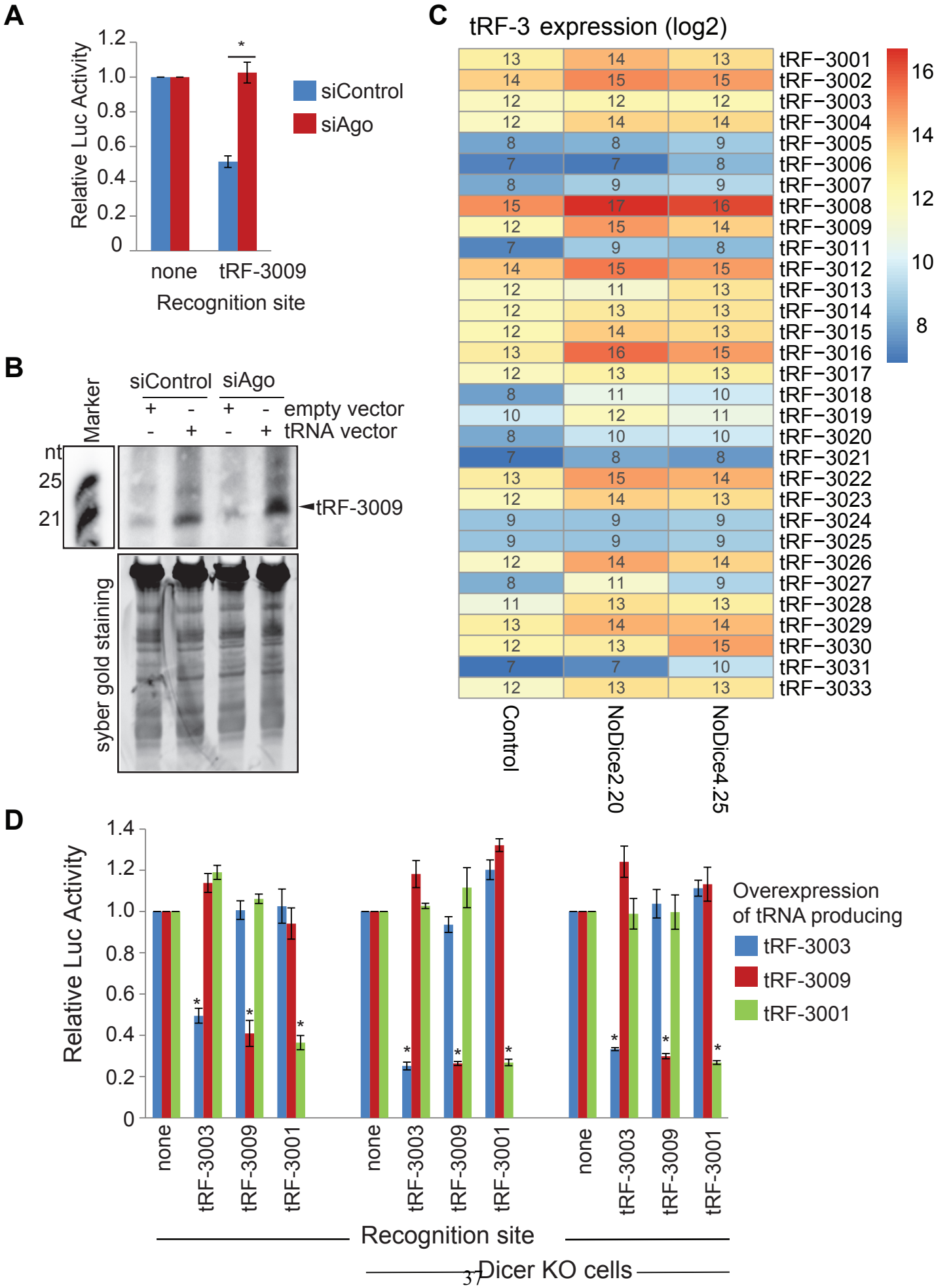

C
tRF-3001
tRF-3003
tRF-3009
empty vector
tRNA-LeuAAG
empty vector
tRNA-CysGCA
empty vector
tRNA-LeuTAA
23nt
21nt
17nt
tRF-3b
tRF-3a
Northern blotting
sybr gold staining

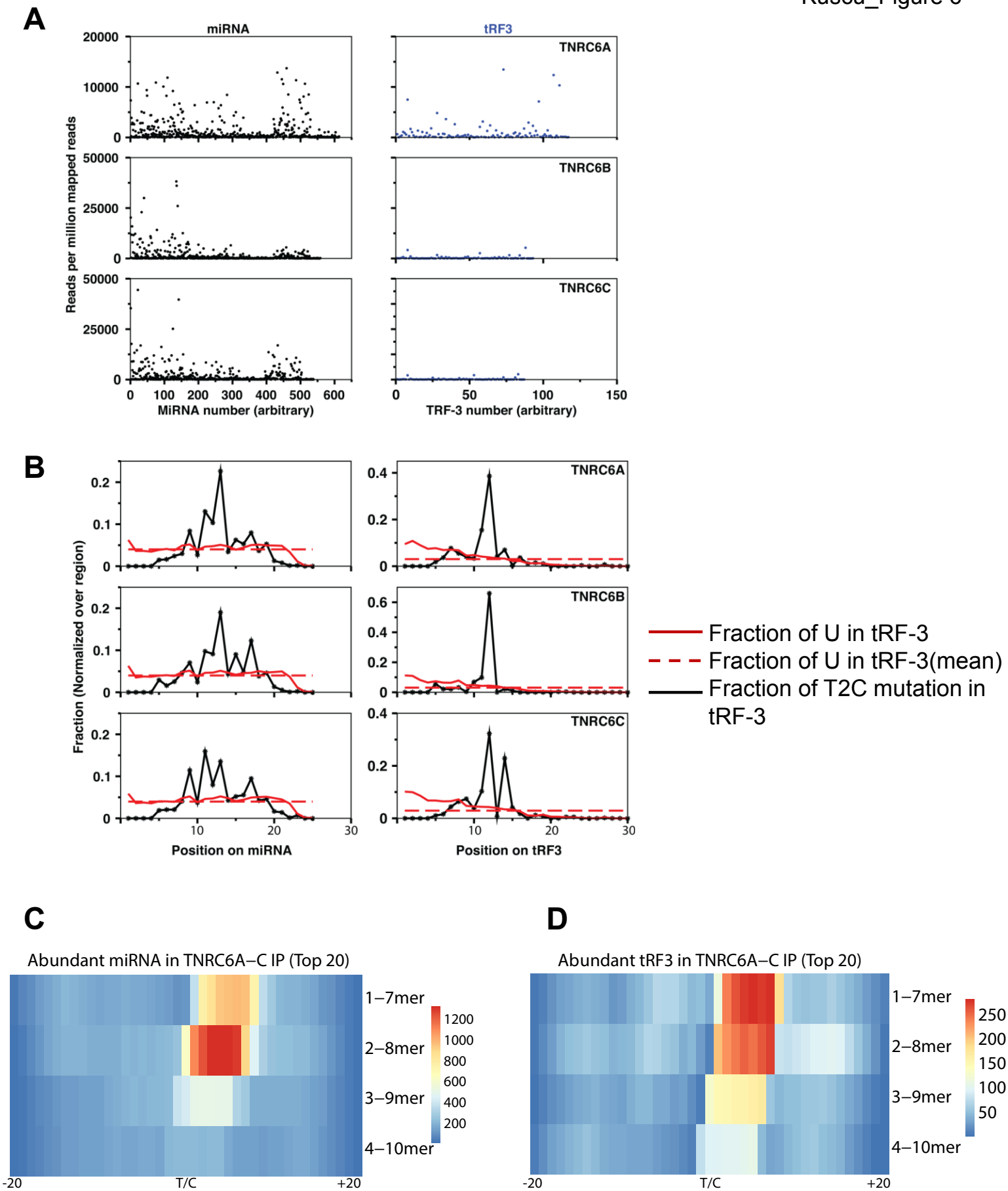

D
pull-down
input
IgG
Ago
Northern blotting
23nt
21nt
17nt
tRF-3009a

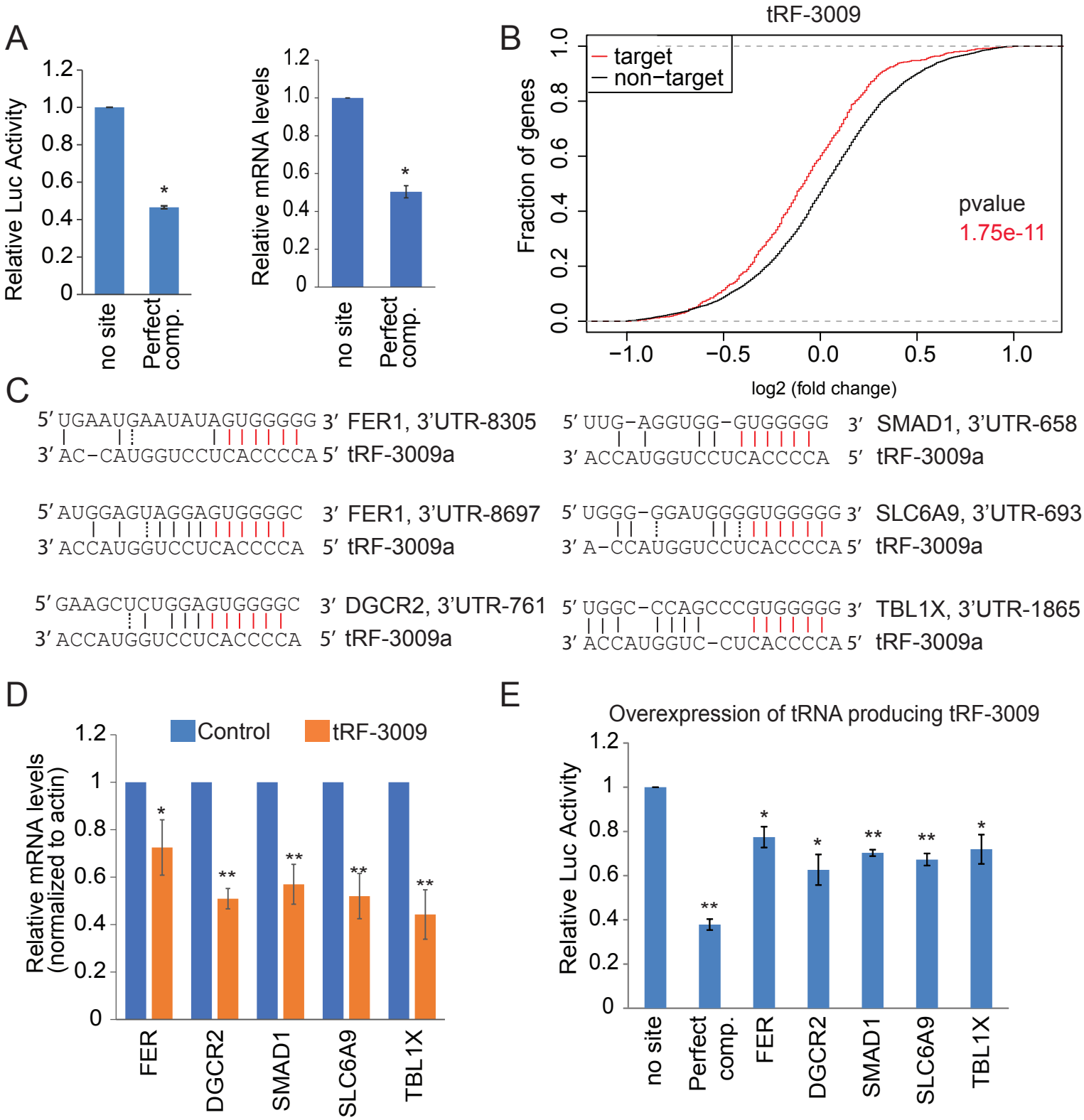
Kuscu_Figure 3



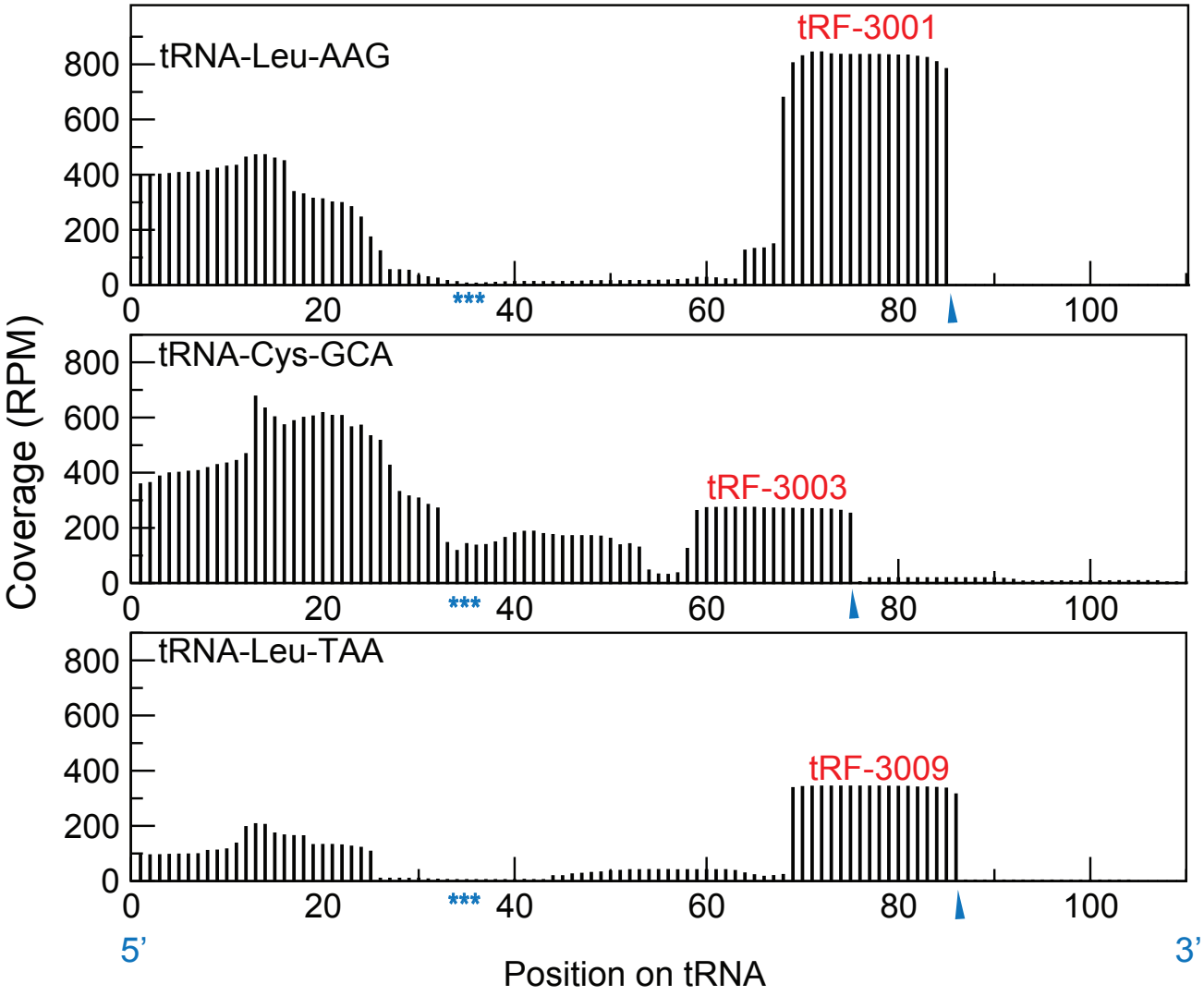

A
Relative Luc Activity
1.2
1
0.8
0.6
0.4
0.2
0
no site
perfect comp
M1
M2
M3
M4
M5
M6
M7
tRF-3001a 3′ ACCACCGTCGCCACCCTA 5′
Perfect comp 5′ TGGTGGCAGCGGTGGGAT 3′
M1 TGGTGGCAGCGGTGGCTA
M2 TGGTGGCAGCGGACCGAT
M3 TGGTGGCAGGCCTGGGAT
M4 TGGTGGGTCCGGTGGGAT
M5 TGGACCCAGCGGTGGGAT
M6 ACCTGGCAGCGGTGGGAT
M7 TGGTGGCAGCGCACCCTT
B
Relative Luc Activity
1.4
1.2
1
0.8
0.6
0.4
0.2
0
no site
perfect comp
M1
M2
M3
M4
M5
M6
M7
M8
M9
tRF-3003a 3′ ACCTCCCCCGTGGGCCT 5′
Perfect comp 5′ TGGAGGGGGCACCCGGATTTGA 3′
M1 TGGAGGGGGCACCCGGATTACT
M2 TGGAGGGGGCACCCGGTAATGA
M3 TGGAGGGGGCACCGCCATTTGA
M4 TGGAGGGGGCTGGCGGATTTGA
M5 TGGAGGGCCGACCCGGATTTGA
M6 TGGACCCGGCACCCGGATTTGA
M7 ACCTGGGGGCACCCGGATTTGA
M8 TGGAGGGGGCACCCGCTAAACA
M9 TGGAGGGGGCTGGGCCATTTGA
C
Relative Luc Activity
1.4
1.2
1
0.8
0.6
0.4
0.2
0
no site
perfect comp
M1
M2
M3
M4
M5
M6
M7
M8
M9
tRF-3009a 3′ ACCATGGTCCTCACCCCA 5′
Perfect comp 5′ TGGTACCAGGAGTGGGGTTCGA 3′
M1 TGGTACCAGGAGTGGGGTTGCT
M2 TGGTACCAGGAGTGGGCAACGA
M3 TGGTACCAGGAGTCCCGTTCGA
M4 TGGTACCAGGTCAGGGGTTCGA
M5 TGGTACCTCCAGTGGGGTTCGA
M6 TGGTTGGAGGAGTGGGGTTCGA
M7 ACCAACCAGGAGTGGGGTTCGA
M8 TGGTACCAGGAGTGGCCAAGCA
M9 TGGTACCAGGACACCCCTTCGA



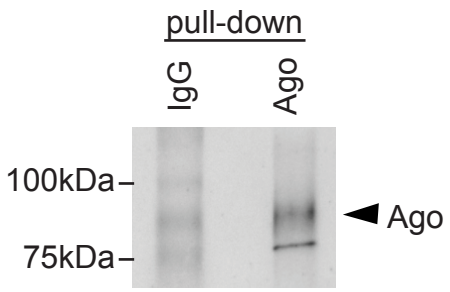

A
miRNA
tRF3
TNRC6A
TNRC6B
TNRC6C
Reads per million mapped reads
MiRNA number (arbitrary)
TRF-3 number (arbitrary)
B
Fraction (Normalized over region)
Position on miRNA
Position on tRF3
Fraction of U in tRF-3
Fraction of U in tRF-3(mean)
Fraction of T2C mutation in tRF-3
C
Abundant miRNA in TNRC6A–C IP (Top 20)
1–7mer
2–8mer
3–9mer
4–10mer
T/C
D
Abundant tRF3 in TNRC6A–C IP (Top 20)



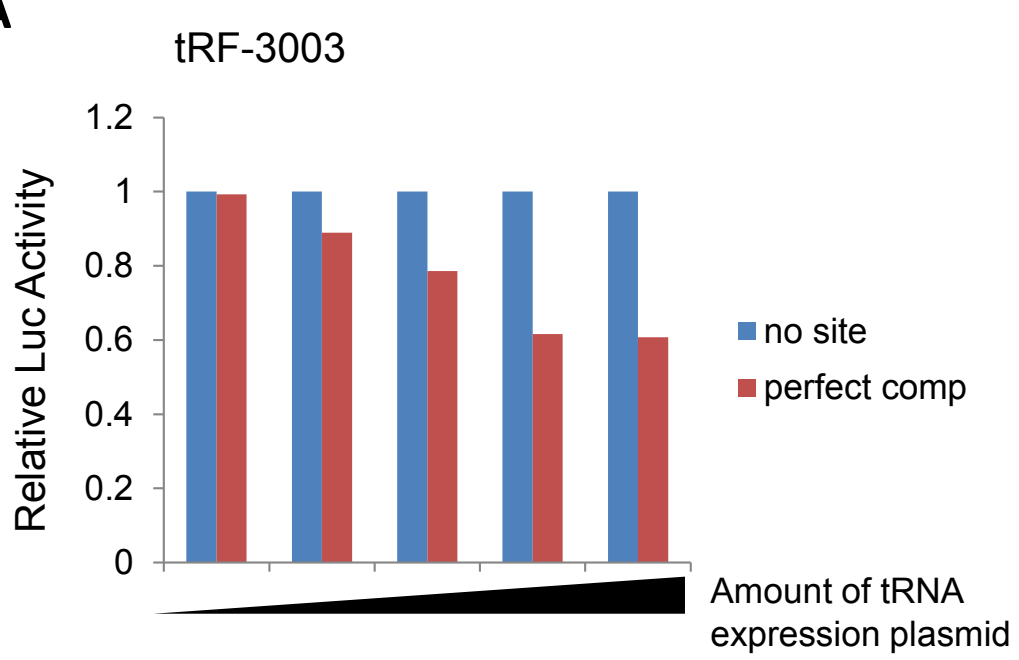
Kuscu_Supp. Fig. S1



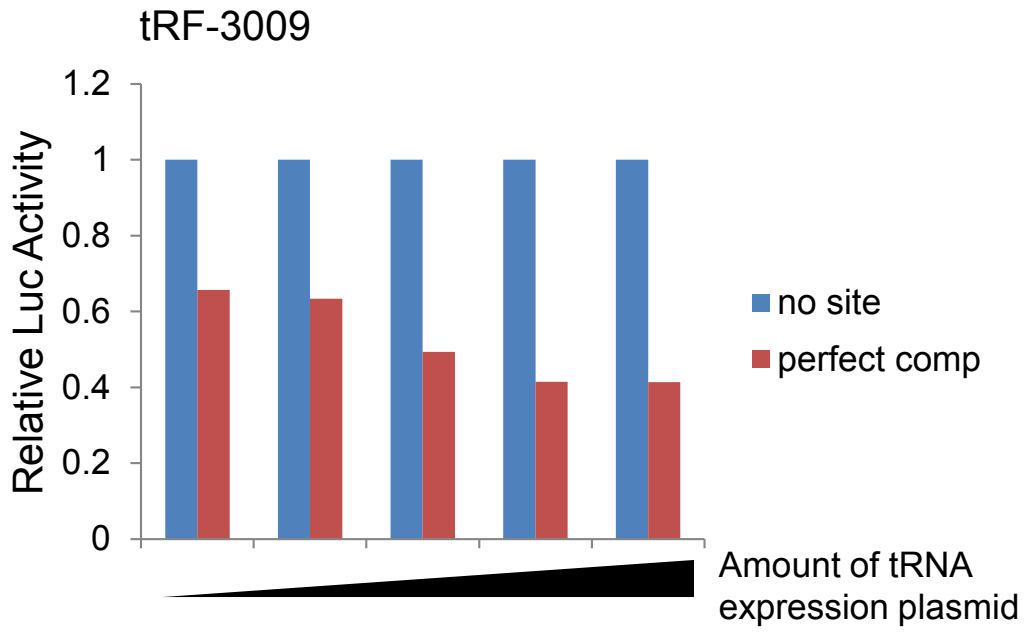
Kuscu_Supplementary Figure S2



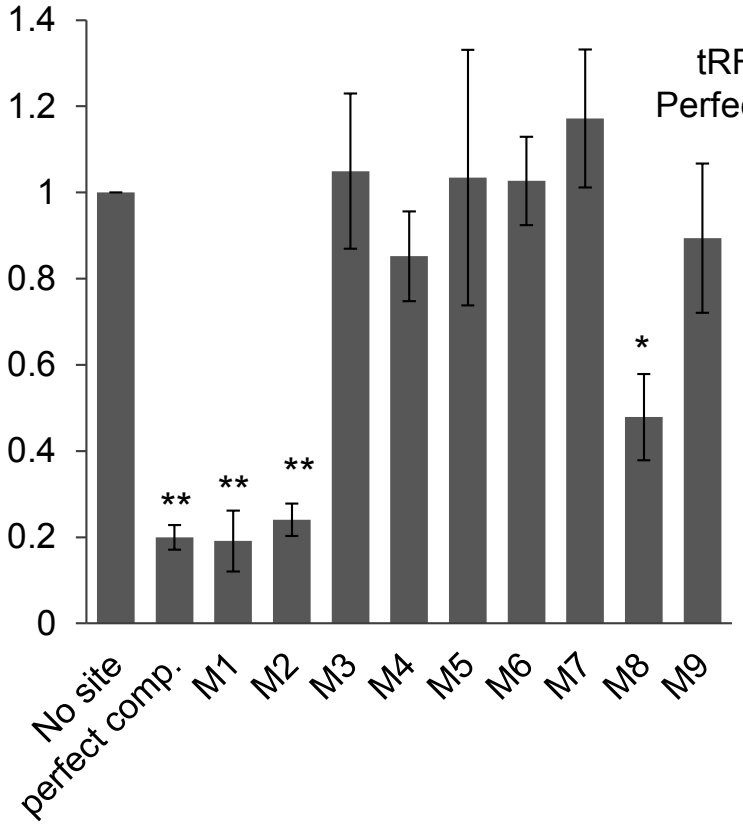
A


tRF-3003
Relative Luc Activity
1.2
1
0.8
0.6
0.4
0.2
0
no site
perfect comp
Amount of tRNA
expression plasmid

B

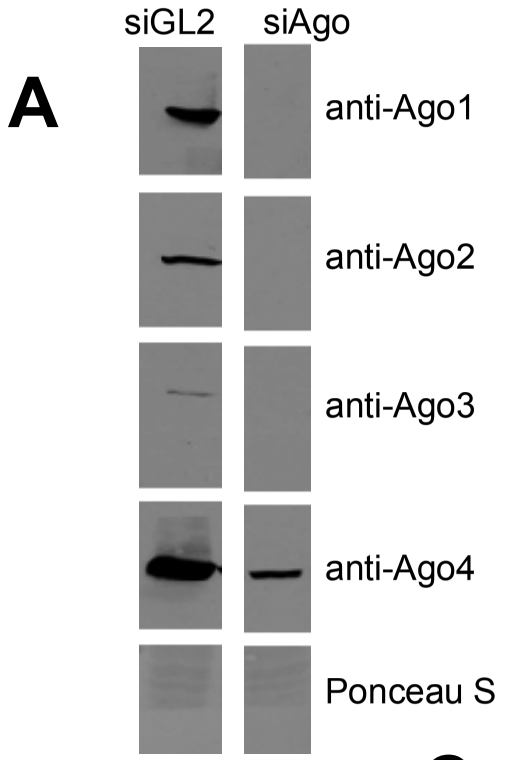

tRF-3009
Relative Luc Activity
1.2
1
0.8
0.6
0.4
0.2
0
no site
perfect comp
Amount of tRNA
expression plasmid

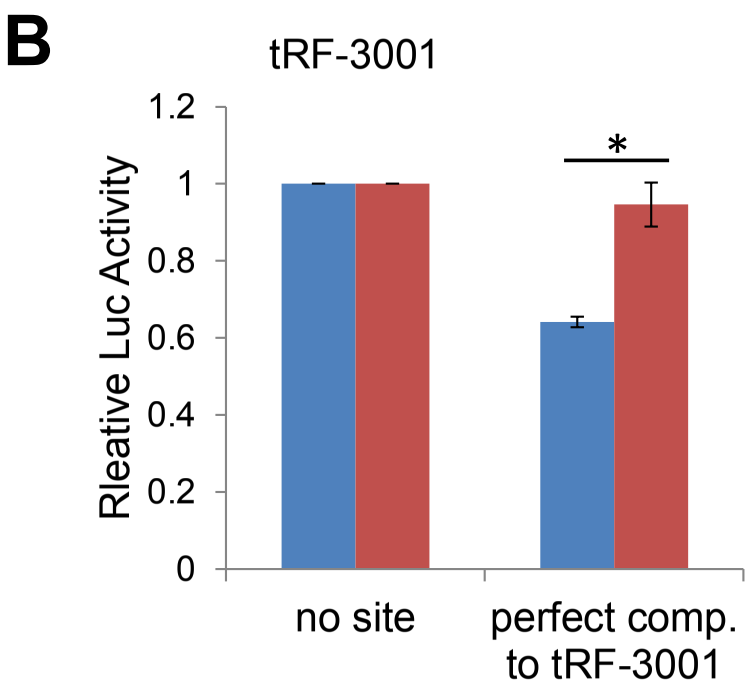
Kuscu_Supp. Fig.S4

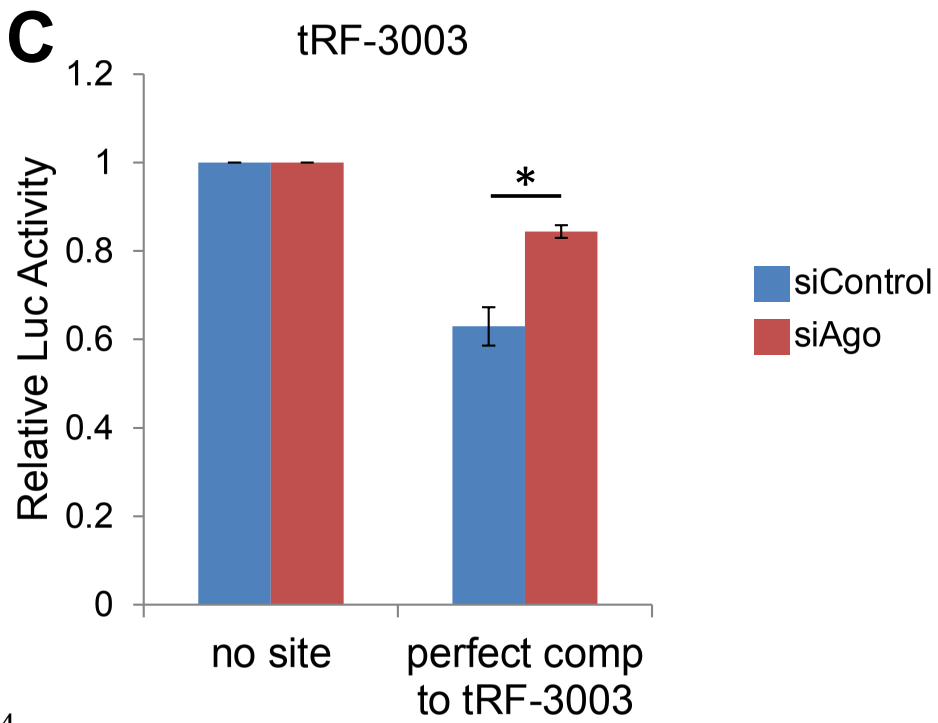


tRF-3009a	3′ **ACCATGGTCCTCACCCCA** 5′
Perfect comp	5′ TGGTACCAGGAGTGGGGTTCGA 3′
M1	TGGTACCAGGAGTGGGGTT**GCT**
M2	TGGTACCAGGAGTGGG**CAA**CGA
M3	TGGTACCAGGAGT**CCC**GTTCGA
M4	TGGTACCAGG**TCA**GGGGTTCGA
M5	TGGTACC**TCC**AGTGGGGTTCGA
M6	TGGT**TGG**AGGAGTGGGGTTCGA
M7	**ACCA**ACCAGGAGTGGGGTTCGA
M8	TGGTACCAGGAGTGG**CCAAGC**A
M9	TGGTACCAGGA**CACCCC**TTCGA

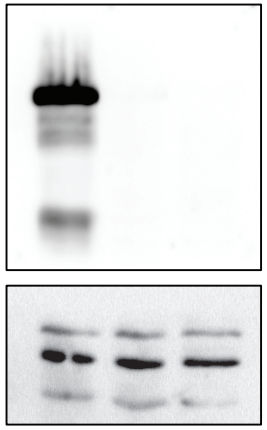
Kuscu_Supp.Fig.S5



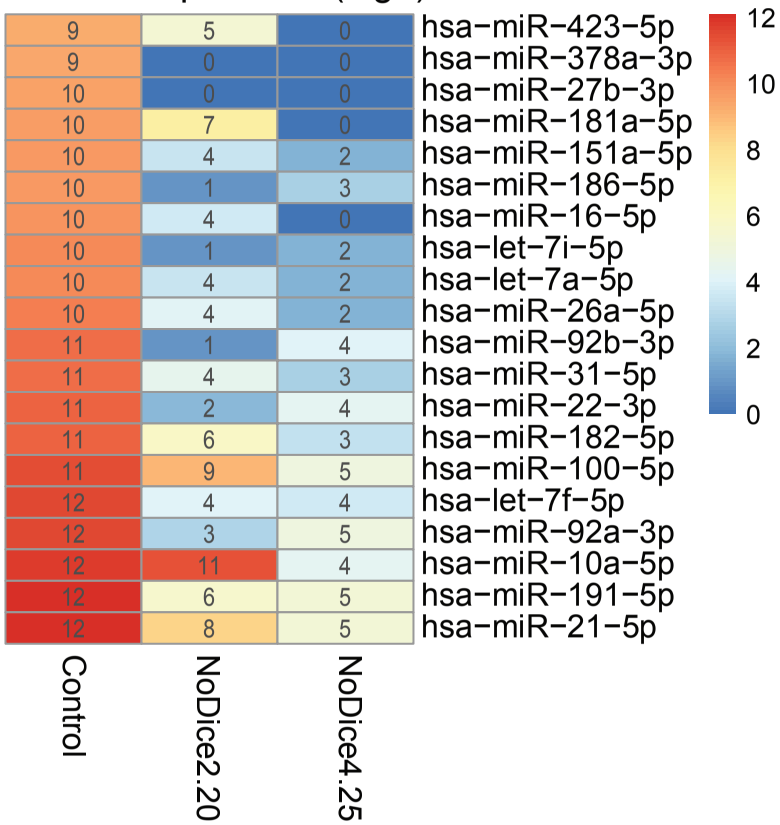




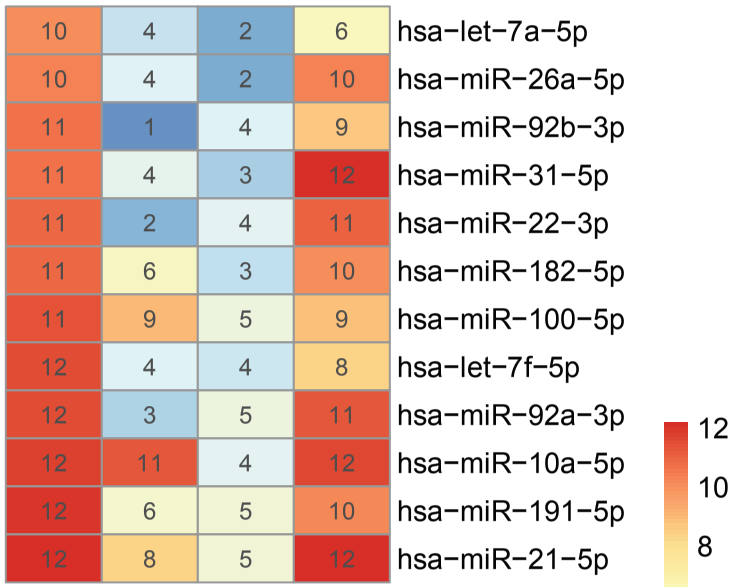
A


Dicer KO
WT
2-20
4-25
α-Dicer
α-tubulin

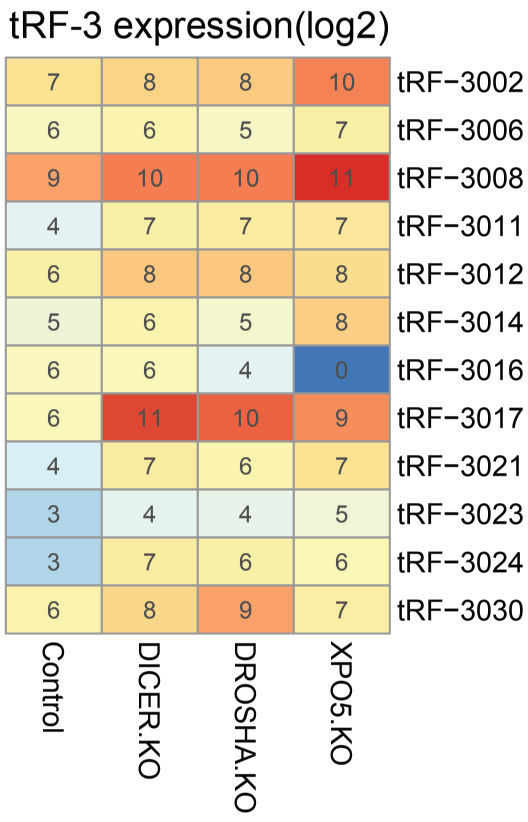
B


miRNA expression (log2)
hsa−miR−423−5p 9 5 0
hsa−miR−378a−3p 9 0 0
hsa−miR−27b−3p 10 0 0
hsa−miR−181a−5p 10 7 0
hsa−miR−151a−5p 10 4 2
hsa−miR−186−5p 10 1 3
hsa−miR−16−5p 10 4 0
hsa−let−7i−5p 10 1 2
hsa−let−7a−5p 10 4 2
hsa−miR−26a−5p 10 4 2
hsa−miR−92b−3p 11 1 4
hsa−miR−31−5p 11 4 3
hsa−miR−22−3p 11 2 4
hsa−miR−182−5p 11 6 3
hsa−miR−100−5p 11 9 5
hsa−let−7f−5p 12 4 4
hsa−miR−92a−3p 12 3 5
hsa−miR−10a−5p 12 11 4
hsa−miR−191−5p 12 6 5
hsa−miR−21−5p 12 8 5
Control
NoDice2.20
NoDice4.25
12
10
8
6
4
2
0

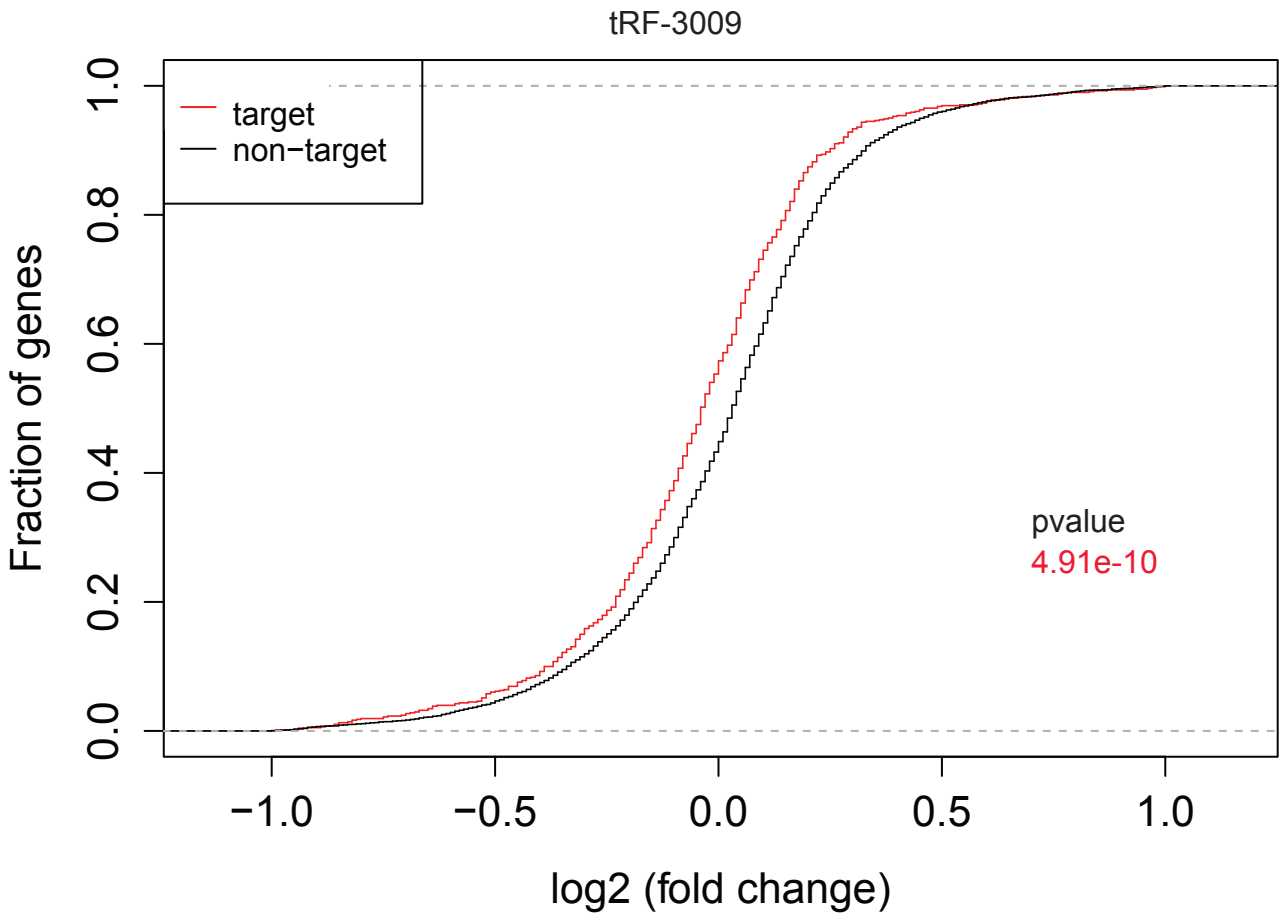
C


miRNA expression(log2)
hsa−let−7a−5p 10 4 2 6
hsa−miR−26a−5p 10 4 2 10
hsa−miR−92b−3p 11 1 4 9
hsa−miR−31−5p 11 4 3 12
hsa−miR−22−3p 11 2 4 11
hsa−miR−182−5p 11 6 3 10
hsa−miR−100−5p 11 9 5 9
hsa−let−7f−5p 12 4 4 8
hsa−miR−92a−3p 12 3 5 11
hsa−miR−10a−5p 12 11 4 12
hsa−miR−191−5p 12 6 5 10
hsa−miR−21−5p 12 8 5 12
12
10
8
6
4
2
0

D


tRF-3 expression(log2)
tRF−3002 7 8 8 10
tRF−3006 6 6 5 7
tRF−3008 9 10 10 11
tRF−3011 4 7 7 7
tRF−3012 6 8 8 8
tRF−3014 5 6 5 8
tRF−3016 6 6 4 0
tRF−3017 6 11 10 9
tRF−3021 4 7 6 7
tRF−3023 3 4 4 5
tRF−3024 3 7 6 6
tRF−3030 6 8 9 7
Control
DICER.KO
DROSHA.KO
XPO5.KO

Kuscu_Supp. Fig.S7



Supplementary Table 1: List of primers used in this study.

primer name	sequence	purpose
pcDNA3_F	CTCGAGCATGCATCTAGAGGG	cloning primer
pcDNA3_R	GGATCCGAGCTCGGTACC	cloning primer
ch17-tRNA26_F	ACCGAGCTCGGATCCCCAAGATTCCTTAATCTAGTTGGTT	cloning primer
ch17-tRNA26_R	CTCTAGATGCATGCTCGAGCTCACTCGCATTGCATTCTAC	cloning primer
Ch6-tRNA83_F	CTTGGTACCGAGCTCGGATCCTGCAGTTGTCTTCACTGCC	cloning primer
Ch6-tRNA83_R	CCCTCTAGATGCATGCTCGAGAAGCAGGACTTTGCGTGTG	cloning primer
Ch16-tRNA16_F	TTGGTACCGAGCTCGGATCCAATCCGGGTCGTATGGATTA	cloning primer
Ch16-tRNA16_F	CCCTCTAGATGCATGCTCGAGCCAGCGTTTCCTCTTACCC	cloning primer
psiCHECK-2_PmeI-F	GTTTAAACCTAGAGCGGCCG	cloning primer
psiCHECK2_XhoI-R	CTCGAGCGATCGCCTAGAAT	cloning primer
FER-10451_F	AGGCGATCGCTCGAGCACAGACAAAGGGGAACTGG	cloning primer
FER-11690_R	GCTCTAGGTTTAAACAGGTTCTGCAGACACATGAGTG	cloning primer
DGCR2-1901_F	AGGCGATCGCTCGAGGCCTGTACCCCAACGGTCT	cloning primer
DGCR2-4475_R	GCTCTAGGTTTAAACCCTCTTCCGGAACACAAGTTT	cloning primer
SMAD1-1824_F	AGGCGATCGCTCGAGGGCATCTGCCTCTGGAAAA	cloning primer
SMAD1-2966_R	GCTCTAGGTTTAAACCGAGAGCATAAGTGAATACAAAAGA	cloning primer
SLC6A9-2321_F	AGGCGATCGCTCGAGTCATTCATGCTCATGTCCCC	cloning primer
SLC6A9-3159_R	GCTCTAGGTTTAAACGGCGCACCGTTATTGCTAC	cloning primer

TBLX1-2131_F	AGGCGATCGCTCGAGAATTCTAATGACC AGCCGTGAA	cloning primer
TBLX1-5596_R	GCTCTAGGTTTAAACCTGGAACACACAC CAGATTGC	cloning primer
3003_F	TCCGGGTGCCCCCTC	qPCR primer
3009_F	ACCCCACTCCTGGTACCA	qPCR primer
3001_F	ATCCCACCGCTGCCAC	qPCR primer
miR-21_F	TAGCTTATCAGACTGATGTTGA	qPCR primer
5016_F	GGGGGTATAGCTCAGTGGTAGAG	qPCR primer
P1-2-3_tRNA-CysGCA_R	AGGGGGCACCCGGATT	qPCR primer
P4_tRNA-LeuTAA_F	ACCAGGATGGCCGAGTG	qPCR primer
P4_tRNA-LeuTAA_R	TACCAGGAGTGGGGTTCGAA	qPCR primer
5019_F	GGTAGCGTGGCCGAGC	qPCR primer
P8-9-10_tRNA_Leu_R	TGGCAGCGGTGGGATT	qPCR primer
3007b_F	TCAATTCTCGCTGGGGCCT	qPCR primer
3006b_F	TCAAGTCCCTGTTCGGGC	qPCR primer
3002b_F	TCAAATCCCGGACGAGCC	qPCR primer
3003b_F	TCAAATCCGGGTGCCCC	qPCR primer
FER-460_F	ATGTCAGCAACGTATCCAAGG	qPCR primer
FER-580_R	GAGCTGTGCCCCTTTCAAC	qPCR primer
DGCR2-432_F	GACGAAGCCAACTGTCCAGA	qPCR primer
DGCR2-551_R	GTTCACCGCGTGGAAGTG	qPCR primer
SMAD1-1013_F	CAGCAGCACCTACCCTCACT	qPCR primer
SMAD1-1144_R	GAGAGCCATCCTGGGTCAT	qPCR primer
SLC6A9-274_F	TGGTAGGAAAAGGTGCCAAA	qPCR primer
SLC6A9-401_R	ATAGCCCACGCTCGTCAGTA	qPCR primer
TBL1X-283_F	GGAAGCCTGCTGGTCCAC	qPCR primer
TBL1X-407_R	GTGGCAGCACGATGAAGAG	qPCR primer

Supplementary Table 2: List of plasmids used in this study.

Plasmid ID	Plasmid name
P1	pcDNA3-CysGCA Chr17-tRNA26
P4	pcDNA3-LeuTAA Chr6-tRNA83
P9	pcDNA3-LeuAAG Chr16-tRNA16
P45	psi-CHECK2-tRF3003exactcomp
P46	psi-CHECK2-tRF3009exactcomp
P55	psi-CHECK2-tRF3003-m1
P56	psi-CHECK2-tRF3003-m2
P57	psi-CHECK2-tRF3003-m3
P58	psi-CHECK2-tRF3003-m4
P59	psi-CHECK2-tRF3003-m5
P60	psi-CHECK2-tRF3003-m6
P61	psi-CHECK2-tRF3003-m7
P62	psi-CHECK2-tRF3003-m8
P63	psi-CHECK2-tRF3003-m9
P64	psi-CHECK2-tRF3009-m1
P65	psi-CHECK2-tRF3009-m2
P66	psi-CHECK2-tRF3009-m3
P67	psi-CHECK2-tRF3009-m4
P68	psi-CHECK2-tRF3009-m5
P69	psi-CHECK2-tRF3009-m6
P70	psi-CHECK2-tRF3009-m7
P71	psi-CHECK2-tRF3009-m8
P72	psi-CHECK2-tRF3009-m9
P74	psi-CHECK2-tRF3001exactcomp
P76	psi-CHECK2-tRF3001-m1
P77	psi-CHECK2-tRF3001-m2
P78	psi-CHECK2-tRF3001-m3
P79	psi-CHECK2-tRF3001-m4
P80	psi-CHECK2-tRF3001-m5
P81	psi-CHECK2-tRF3001-m6
P82	psi-CHECK2-tRF3001-m7
P101	psiCHECK2-FER-10451-11690
P104	psi-CHECK2-DGCR2-1901-4475
P105	psi-CHECK2-SMAD1-1824-2966
P106	psi-CHECK2-SLC6A9-2321-3159
P108	psi-CHECK2-TBLX1-2131-5596

RNA
A PUBLICATION OF THE RNA SOCIETY

tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer independent manner

Canan Kuscu, Pankaj Kumar, Manjari Kiran, et al.

RNA published online May 29, 2018

Molecular Neurobiology
https://doi.org/10.1007/s12035-018-1416-y

A Prognostic Signature for Lower Grade Gliomas Based on Expression of Long Non-Coding RNAs

Manjari Kiran[1] · Ajay Chatrath[1] · Xiwei Tang[2] · Daniel Macrae Keenan[2] · Anindya Dutta[1]

Received: 5 July 2018 / Accepted: 25 October 2018

Abstract

Diffuse low-grade and intermediate-grade gliomas (together known as lower grade gliomas, WHO grade II and III) develop in the supporting glial cells of brain and are the most common types of primary brain tumor. Despite a better prognosis for lower grade gliomas, 70% of patients undergo high-grade transformation within 10 years, stressing the importance of better prognosis. Long non-coding RNAs (lncRNAs) are gaining attention as potential biomarkers for cancer diagnosis and prognosis. We have developed a computational model, UVA8, for prognosis of lower grade gliomas by combining lncRNA expression, Cox regression, and L1-LASSO penalization. The model was trained on a subset of patients in TCGA. Patients in TCGA, as well as a completely independent validation set (CGGA) could be dichotomized based on their risk score, a linear combination of the level of each prognostic lncRNA weighted by its multivariable Cox regression coefficient. UVA8 is an independent predictor of survival and outperforms standard epidemiological approaches and previous published lncRNA-based predictors as a survival model. Guilt-by-association studies of the lncRNAs in UVA8, all of which predict good outcome, suggest they have a role in suppressing interferon-stimulated response and epithelial to mesenchymal transition. The expression levels of eight lncRNAs can be combined to produce a prognostic tool applicable to diverse populations of glioma patients. The 8 lncRNA (UVA8) based score can identify grade II and grade III glioma patients with poor outcome, and thus identify patients who should receive more aggressive therapy at the outset.

Keywords Long non-coding RNAs · Gliomas · Gene expression profiling · Prognosis

Abbreviations

lncRNA	Long non-coding RNAs
WHO	World Health Organization
LGG	Lower grade gliomas
GBM	Glioblastoma multiforme
CNS	Central nervous system
TCGA	The Cancer Genome Atlas
CGGA	Chinese Glioma Genome Atlas
HR	Hazard ratio
PFS	Progression-free survival
IFNG	Interferon gamma
Cindex	Concordance index
AUC	Area under curve
ROC	Receiver operating characteristics
UVA8	University of Virginia 8
L1-LASSO	L1 least absolute shrinkage and selection operator
MGMT	O6-methylguanine DNA methyltransferase
FPKM	Fragment per kilobase per million
GTF	Gene transfer format

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12035-018-1416-y) contains supplementary material, which is available to authorized users.

✉ Anindya Dutta
ad8q@virginia.edu

[1] Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Pinn Hall 1232, Charlottesville, VA 22908, USA

[2] Department of Statistics, University of Virginia, Charlottesville, VA 22904, USA

Introduction

Over the past decade, high-throughput RNA-seq technology discovered many novel transcriptional units, which were otherwise missed by probe design based transcriptome profiling. Among these transcriptional units were many long non-coding RNAs (lncRNA), which are transcripts longer than 200 bases with almost no protein-coding potential or open reading frames of < 50 amino acids. These lncRNAs are

Published online: 03 November 2018

numerous in cells [1], are highly regulated, and are more cell-type specific than protein-coding genes [2]. LncRNAs are involved in a broad spectrum of function and recent studies suggest they have specific roles in different diseases like cancer (reviewed in [3, 4]).

Gliomas are the most common form of primary malignant brain tumor, which originate in the supporting glial cells in the brain, including astrocytes, oligodendrocytes, and ependymal cells. Based on WHO 2016 grading system, gliomas are classified into lower grade and much aggressive high-grade gliomas. Grade I is mostly benign, whereas diffuse low-grade and intermediate-grade gliomas make up the WHO grade II and III lesions. Grade IV gliomas include secondary glioblastomas (derived from lower grade gliomas) and primary glioblastoma multiforme (GBM). Surgical resection of tumor is the most common initial treatment for gliomas followed by radiation therapy and chemotherapy, which can increase survival to 12 months [5, 6]. Molecular markers like 1p/19q co-deletion, MGMT promoter methylation, and mutation in IDH1 gene are strong predictors of survival for gliomas [7]. Lower grade gliomas have a better prognosis than high-grade gliomas. Despite a better prognosis for lower grade gliomas than the grade IV tumors, 70% of patients from the former group undergo high-grade transformation within 10 years.

LncRNAs are widely expressed in the central nervous system (CNS) and are involved in several pathways related to CNS development [8–13]. LncRNA BRN1B is one of the critical lncRNAs for brain development [13]. LncRNA Sox2OT plays an important role in determining neural fate [14]. Dysregulation of many lncRNAs like DGCR5, NRON, H19, and DISC2 have been associated with different CNS diseases [15–18]. Previous studies have shown that specific lncRNA expression patterns are also associated with different histological subtypes and grade in gliomas [19, 20]. For example, expression of MALAT1, POU3F3, and H19 are highly correlated with glioma malignancy. More recently, lncRNAs are also found to be of prognostic significance suggesting their role in glioma malignancies and as a potential therapeutic target and biomarker [19, 20]. Li et al. revealed three molecular subtypes of gliomas based on lncRNAs expression that has a strong correlation with patient's survival [21]. Furthermore, analysis on previously published microarray data has explored lncRNA-based signature as a prognostic marker in gliomas ([20, 22–25]).

Many studies have highlighted the power of gene expression profiles to predict tumor classification, patient outcome, and tumor response to therapy. Differentially expressed genes in cancer patients versus normal individuals are often the starting set to predict prognostic signature associated with survival [26–28]. This strategy suffers from false negatives and from the fact that differentially expressed genes might not be associated with differences in survival at all [29]. Another limitation of this method is the requirement of perfect matched normal to identify differentially expressed genes. This creates a major hurdle in case of brain cancer where getting a perfect matched normal tissue is not trivial. While high-throughput technologies have facilitated the search of biomarkers through multivariate data analyses, there still remain challenges with respect to meaningful statistical and biological information. Firstly, most of the biological datasets suffer with multicollinearity: the influence of one gene on expression of other genes. Secondly, there are more features (genes) than observations (patients), which lead to overfitting by most of existing learning algorithms and results in poor performance of the model in prediction in an unseen testing dataset. Thus, a more robust approach is required to find genes as prognostic signature from a multi-dimensional multivariate gene expression data. Regression models like lasso, ridge, and elastic net are some widely used approaches to penalize the effect of multicollinearity and are well suited for constructing models when there are large numbers of features.

In the present study, we develop an lncRNA-based prognostic signature in combination with Cox regression and L1-LASSO regularization to model survival of grade II and grade III glioma patients. This is the first study that combined Cox and lasso regularization to select lncRNAs that can predict survival in glioma patients. After controlling for covariates associated with glioma survival (age, grade, IDH1 mutation status), we selected 8 lncRNAs UVA8, to calculate a risk score, which successfully divides patients into high-risk and low-risk groups in both TCGA (461 patients) and CGGA (274 patients) dataset. The risk score calculated by these eight lncRNAs is an independent and better prognostic marker for grade II and grade III glioma patient survival. The guilt-by-association analysis of lncRNAs in UVA8 indicated their role in suppressing interferon signaling pathway and epithelial to mesenchymal transition. Besides their use as a biomarker, these lncRNAs need to be studied in detail to determine how they affect patient outcome.

Materials and Methods

Patients and Samples

Aligned bam files and clinical information for 512 LGG patients (grade II and III) were retrieved from The Cancer Genome Atlas (TCGA) data portal https://portal.gdc.cancer.gov/. The study is performed on 461 patients for which both RNA-seq and survival information were available. Most samples in TCGA are collected from patients from the USA and also from other countries, including Canada, Russia, and Italy. This dataset being the largest and most updated glioma dataset is used as training dataset in the present study. The raw sequencing data for 274 glioma patients (175 grade II and III) from Chinese Glioma Genome Atlas (CGGA) as independent

cohort was downloaded using accession no. SRP027383 [30]. The survival information for these Chinese patients was downloaded from CGGA http://www.cgga.org.cn/. IDH1 mutation data for all the LGG patients were retrieved from Tier 3 TCGA data accessed from the Broad GDAC Firehose; https://gdac.broadinstitute.org.

RNA-Seq Data Quantification and Analysis

The most recent version of Gencode (GENCODE v 26) GTF file available at the time of this study was used for the gene quantification [31]. Gene abundance in FPKM (fragment per kilobase per million) was obtained for 58,219 genes with 15,787 genes annotated as lncRNA in GENCODE v26 using Stringtie v1.3.3 [32]. Out of 15,787 lncRNAs, 1289 lncRNAs with a median expression of 1 FPKM in 512 LGG patients were finally considered for the survival model.

Survival Model Selection Process

The gene expression data for lncRNAs was *Z*-score transformed to avoid systematic error across different experiments. We first randomly selected 60% of TCGA patients for training set and remaining 40% of TCGA patients for testing set. Since, clinical information like age, gender, tumor grade, or IDH mutation status can have an effect on survival (Fig. S1), we assessed the prognostic potential of each lncRNA by multivariate Cox regression controlling the effects from these other variables. We used FDR corrected p value cutoff of 0.05 obtained after log-likelihood test comparing restricted (age, gender, tumor grade, and IDH mutation status) with unrestricted (lncRNA expression, age, gender, tumor grade, and IDH mutation status) model to identify the significant association of an lncRNA with survival. We used Cox-proportional hazards model based on L1-penalized (LASSO) estimation to select the best model comprising a subset of prognostic lncRNA [33–35]. We used LASSO because it is suited for constructing models when there is a large number of correlated covariates [34].

Risk Score Calculation

Risk score for each patient was established by including each of the selected genes weighted by their estimated regression coefficients in the multivariable Cox regression analysis as discussed in previous studies [36, 37].

UVA8 risk score = (− 0.378 × expression value of RP11-266 K4.14) + (− 0.301 × expression value of FLJ37035) + (− 0.280 × expression value of LINC01561) + (− 0.368 × expression value of RP11-118 K6.3) + (− 0.369 × expression value of DGCR9) + (− 0.299 × expression value of RP11-142A22.3) + (− 0.434 × expression value of LINC00641) + (− 0.543 × expression value of RP11-96H19.1).

Coefficients are median Cox coefficient (after lasso selection and multivariate Cox regression) for each of the eight lncRNAs from the successful models (models which can stratify patients in testing set).

Statistical Analysis

R package glmnet was used to perform L1-penalized cox regression (L1-least absolute shrinkage and selection operator) [38]. R package survival and survminer were used for survival data analysis and generating Kaplan–Meier plots. Different survival models were compared by time-dependent concordance index (Cindex) [39]. Cindex is the most commonly used performance measure for survival models, which calculates the fraction of pairs whose predicted survival time is correctly ordered. R package pec::cindex is used to calculate time-dependent cindex [40].

Results

Building the lncRNA-based Survival Model

We developed an lncRNA-based survival model for gliomas through the following steps (Fig. 1).

1) We first randomly selected 60% ($n = 277$) of the patients from TCGA as training set and reserved the remaining 40% ($n = 184$) of patients as testing set. The results remain similar with 70% patients in training and 30% in testing set (Fig. S3 A).
2) Cox multivariate regression was carried out in the training set on 1289 lncRNA controlling for effects from other covariates like age, gender, tumor grade, and IDH1 mutation status.
3) LncRNAs significantly associated with survival after likelihood ratio test (FDR, $p < 0.05$) were retained for selecting lncRNAs by lasso regularization.
4) After lasso regularization and lncRNA selection, a risk score formula was established by including selected lncRNAs weighted by their estimated regression coefficients in the multivariable Cox regression analysis. Risk score = $\sum_{i=1}^{n} \beta_i * x_i$ (where, β is coefficient and x is expression level of lncRNA i)
5) Patients were classified into high-risk and low-risk group by using the median risk score as the cutoff in the training set. The coefficient for each lncRNA and cutoff of risk score obtained from training set was used to calculate risk score and stratify patients into two groups in testing set.

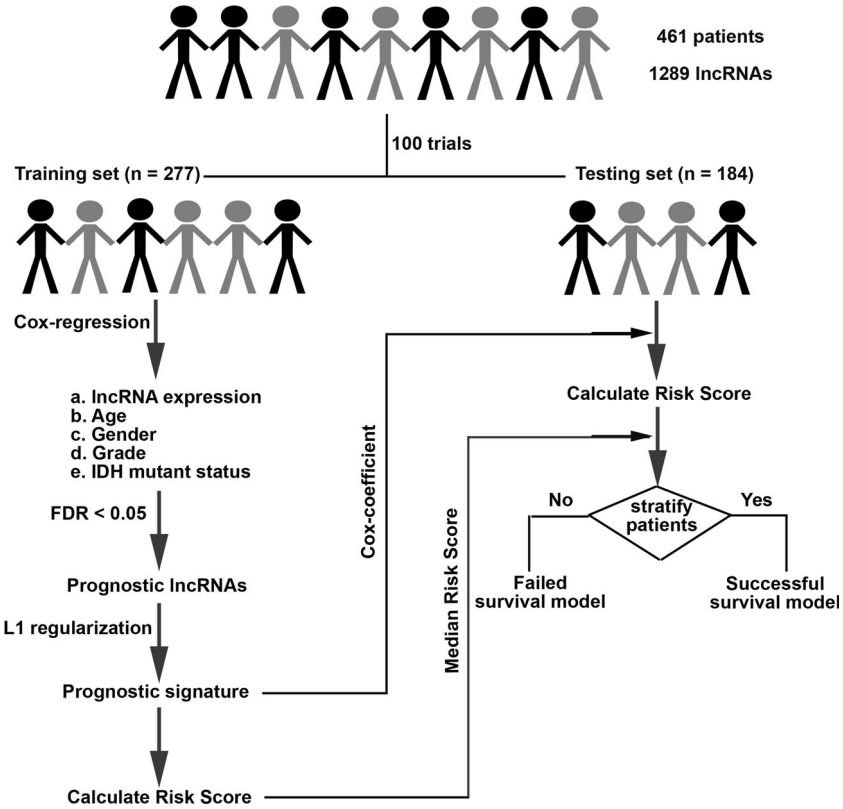


Fig. 1 Flowchart showing steps involved in identification of lncRNA-based prognostic signature

6) Survival differences between the low-risk and high-risk groups in the training and testing sets were assessed by the Kaplan–Meier estimate and compared using the log-rank test.

Steps 1–6 were repeated 100 times to obtain up to 100 different lncRNA subsets (models). Only those models that separated patients in the testing set such that those with low-risk score had significantly better survival than those with high-risk score were considered as successful models and retained.

The result obtained from one such survival model is shown in Fig. S2. In ~20% of the trials the multivariate cox-regression and lasso regularization in the training set did not select any lncRNAs significantly associated with survival (NA in Fig. 2a). The remaining 80% of the survival models contained different numbers of lncRNAs (*x*-axis of Fig. 2a) that significantly stratify patients into low- and high-risk groups in training set (Fig. 2a). Among these 80% of survival models, 86% also significantly separated patients into high risk and low risk in the testing set and are referred to as successful survival models. In order to create a robust survival model we sorted the lncRNAs based on the number of times an lncRNA was selected by successful survival models (Fig. 2b). Out of 167 total prognostic lncRNA in 69 successful survival models, we first ranked lncRNAs based on number of times a given RNA was selected by successful models and then from the top 20 selected 8 lncRNAs with the highest median Cox coefficient (absolute value > 0.2) and least variance in the successful models in the testing set (absolute value < 0.10). Seven out of these 8 lncRNAs were also selected after 70–30% split of training and testing patients (Fig. S3A), after 1000 trials instead of 100 (Fig. S3B), and all 8 lncRNAs were selected when we used elastic net, instead of lasso, for regularization and lncRNA selection (Fig. S3C) suggesting the prognostic importance of these 8 lncRNAs in gliomas. For brevity, this set of eight lncRNAs as a prognostic signature of gliomas will be referred to as UVA8 (University of Virginia 8) in the manuscript. AF131216.5 is associated with high median Cox coefficient (− 0.83) but with high variance (0.35, greater than defined cutoff of 0.10). Despite its high variance, we tested whether including AF131216.5 in the final model will improve the performance of UVA9 on our training (TCGA) and testing dataset (CGGA). The result is described below.

UVA8 is Predictive of Survival in Training and Independent Validation Set

We assessed the predictive power of UVA8 by comparing overall survival of low- and high-risk patients in the entire TCGA dataset stratified based on median risk score obtained by UVA8 (risk score calculation discussed in methods).

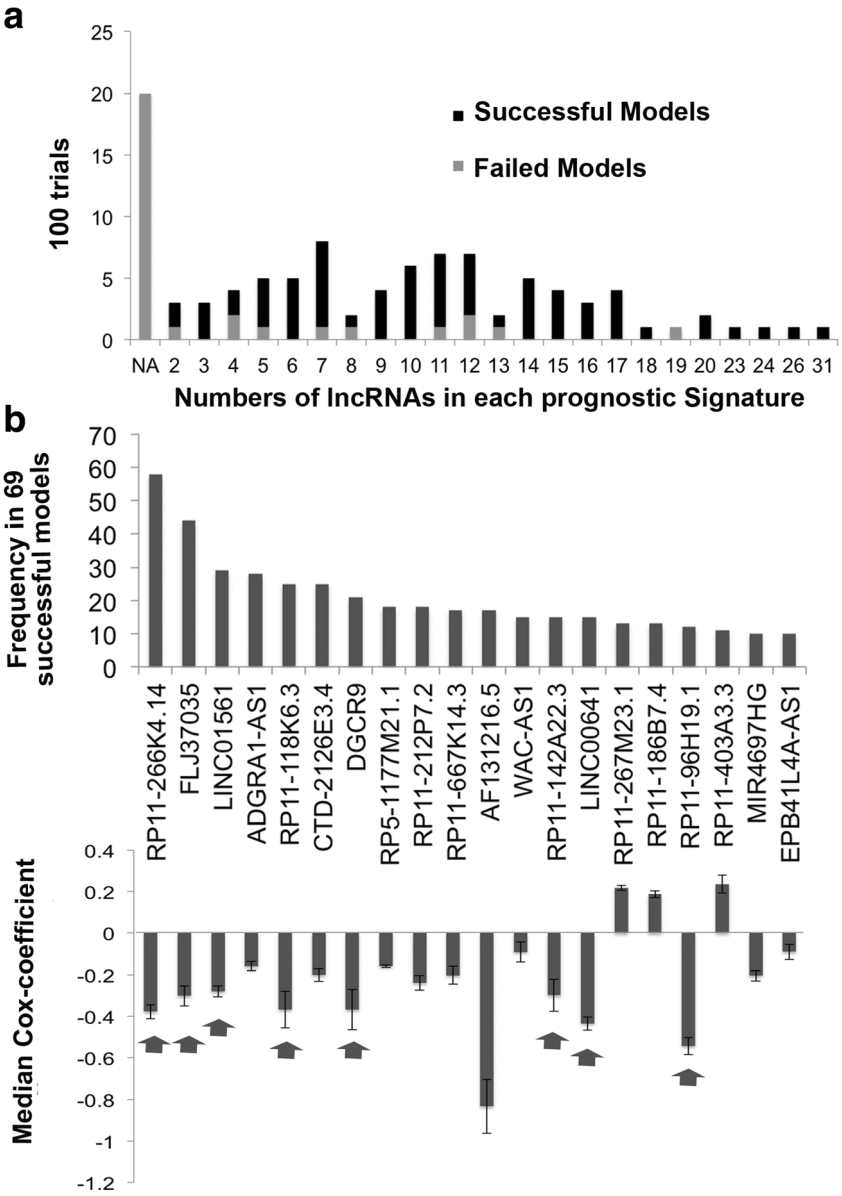


Fig. 2 Selection of lncRNAs with best predictors of outcome. **a** Barplot showing number of lncRNAs that predicted outcome in the training set in 100 trials. The successful models were those that also predicted outcome in the testing set. NA: no lncRNA predicted outcome in training set. **b** Barplot showing number of times each of the top 20 lncRNAs (out of 167) were present in successful survival models (significant in testing set). The lower panel shows median Cox coefficient (after lasso penalization and multivariate Cox regression) and the variance of the Cox coefficient for each of the above 20 lncRNAs from the successful models where they were selected. The arrow points towards lncRNAs selected for UVA8

Patients in the low-risk group showed longer overall survival than the high-risk group in TCGA dataset (Fig. 3a, median OS 741.5 vs. 639 days; $P = 3.1e\text{-}15$, HR = 5.8). The risk scores of the patients in the TCGA dataset range from − 4 to 4 with median risk score of − 0.023 (Fig. 3b, top panel). Moreover, there are more patients alive in the low-risk group than in the high-risk group (Fig. 3b, middle panel). Interestingly, expression levels of all lncRNA in UVA8 are high- in low-risk patients than in high-risk patients indicating these lncRNAs as favorable prognostic genes (Fig. 3b, bottom panel). These findings were further validated in an independent validation dataset comprising of 274 patients obtained from CGGA. Using the same median coefficient of UVA8 obtained from the successful survival models in TCGA, patients showed longer overall survival in low-risk than in high-risk group in CGGA (Fig. 3c, median OS = 1120.5 vs. 587 days; $P = 0.0017$, HR = 1.68). Moreover, low-risk group in CGGA has also longer progression-free survival (PFS) than the high-risk group (Fig. 3d, median PFS 597.5 vs. 411.5 days; $P = 0.00088$, HR = 1.70). Thus, UVA8 can predict survival in both training and independent validation set.

Since, 32% of patients in CGGA are in grade IV, the difference in overall survival could be due to over-representation of grade IV patients in high-risk group. However, even when only lower grade gliomas (grade II and III) were separately examined we found significantly longer survival for low-risk versus high-risk patients (Fig. S4A). UVA8 fails to cluster grade IV patients from CGGA into two distinct groups highlighting the specificity of signature for lower grade gliomas (Fig. S4B). We also assessed the predictive capability of UVA9 (UVA8+ AF131216.5) on TCGA and CGGA and noticed almost no improvement in TCGA (P 3e-15, HR =

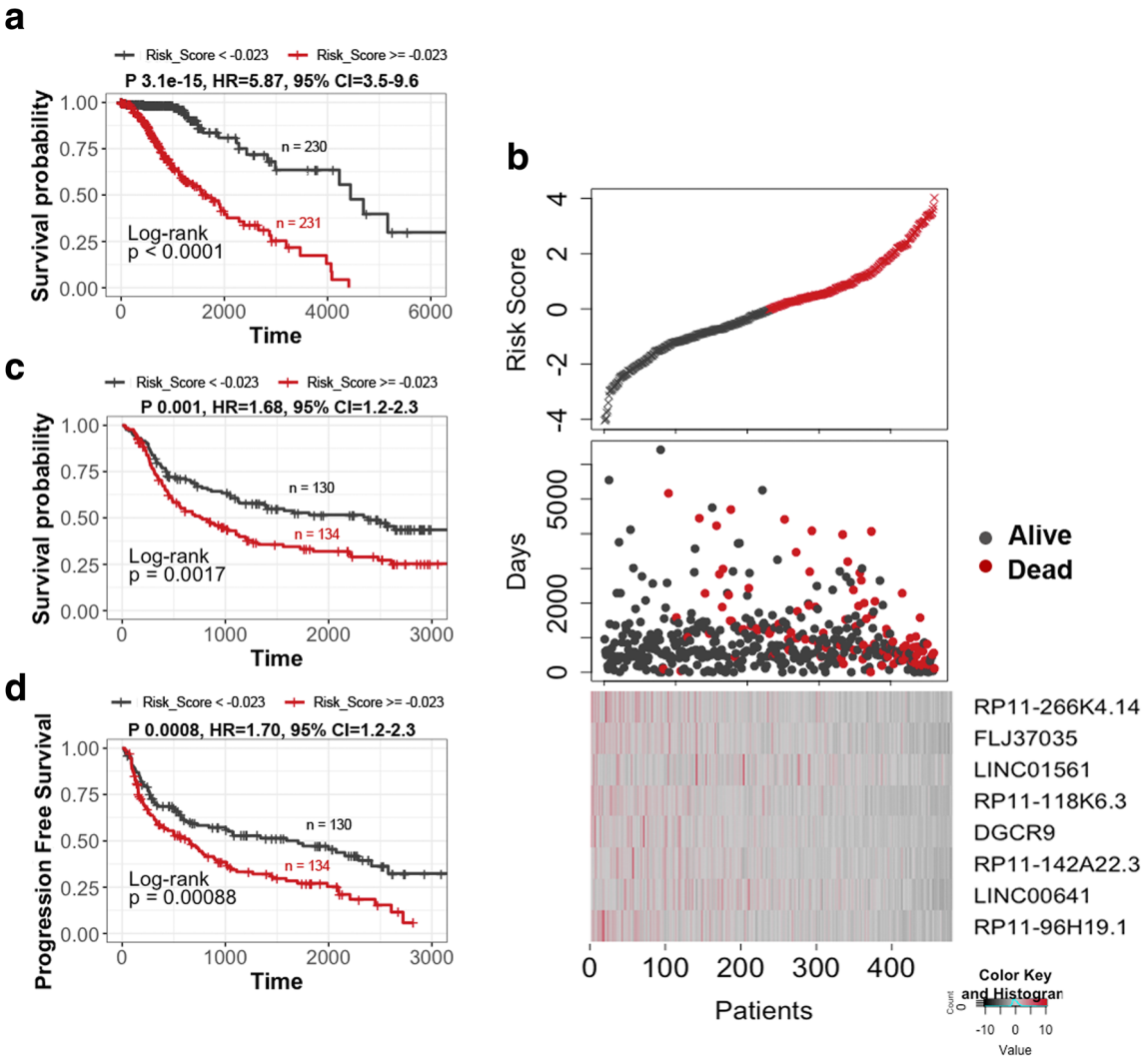
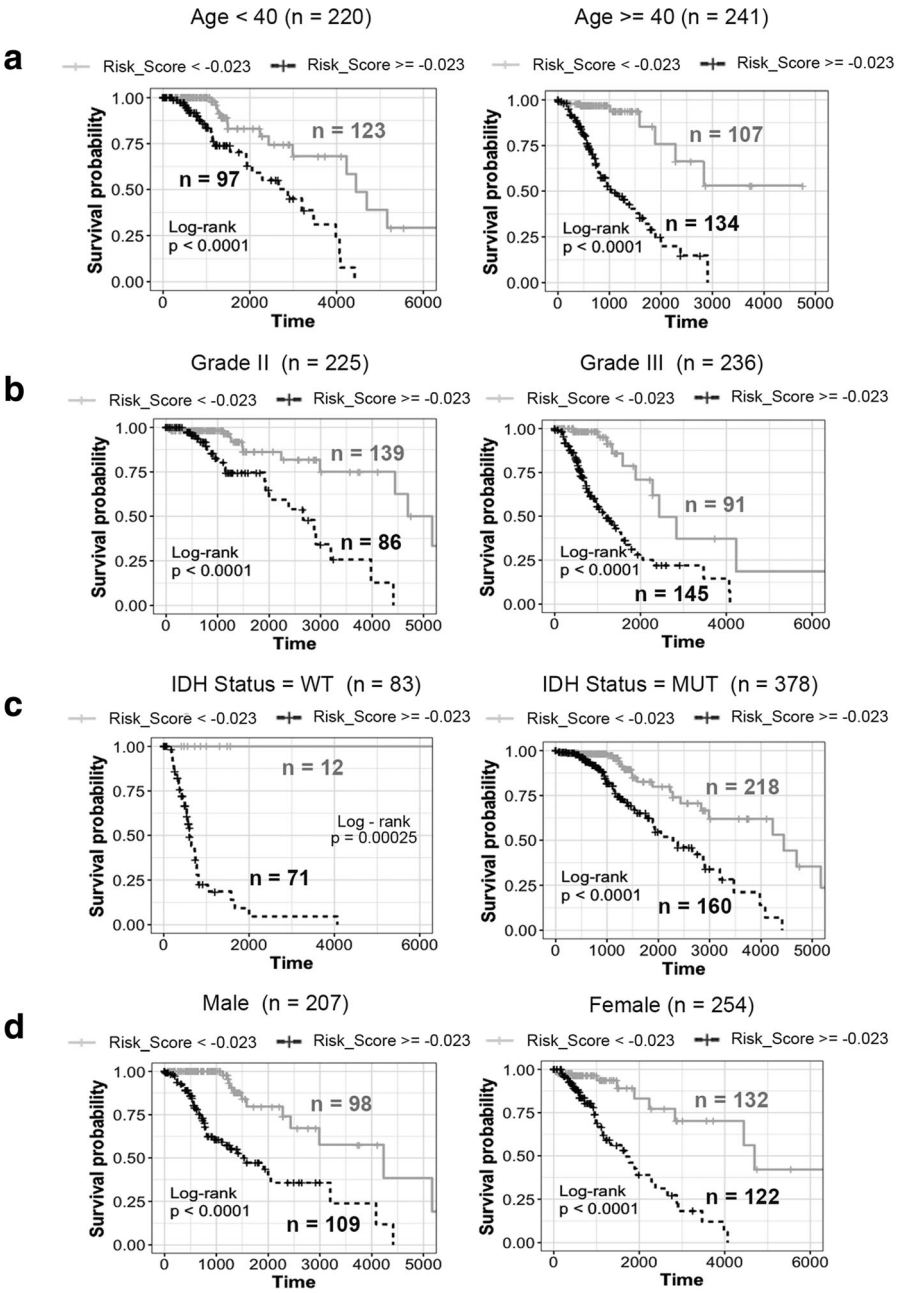


Fig. 3 Survival analysis of the patients divided by the prognostic lncRNAs in two data sets. **a** Patients in the entire TCGA dataset with risk score greater than median score of −0.023 show poor survival compared with patients with risk score less than median risk score. **b** Upper panel: plot showing patients sorted based on UVA8 risk score with black representing patient with risk score below median and red showing those with risk score above median. Middle panel: Number of days of survival indicated on *Y*-axis of patients sorted on the *X*-axis based on the risk scores in the top panel and alive/dead status indicated by color. Bottom panel: *z*-score transformed expression value of lncRNAs in UVA8 show higher expression in patients with low risk score. **c** Kaplan–Meier plot of overall survival of patients in CGGA dataset with risk score greater than (red) or less than (black) median risk score of TCGA dataset. **d** Kaplan–Meier plot for progression-free survival in CGGA dataset showed poor survival for patients with high-risk score. Rest as in (**c**)

5.623.95% CI = 3.47–9.11) and marginal improvement in CGGA (*P* 0.02, HR = 1.74, 95% CI = 1.10–2.77) compared to UVA8 (TCGA *P* 3.1e-15, HR = 5.87, 95% CI = 3.5–9.6, CGGA *P* 0.03, HR = 1.66, 95% CI = 1.05–2.64) (Fig. S5).

8 lncRNA-based Risk Score is an Independent Predictor of Survival

Lower grade gliomas have poorer outcomes in older patients, in tumors of higher grade and tumors with wild-type IDH1 status (Fig. S1). Interestingly, the risk score derived from UVA8 is higher in patients older than 40 years, patients in grade III vs. grade II and patients harboring wild-type IDH1 gene (Fig. S6). It was therefore important to determine whether UVA8-derived risk score is an independent predictor of survival. We divided the patients into younger (age < 40) and older (age ≥ 40) groups and found that risk score can still stratify the patients into low risk and high risk in both groups (Fig. 4a). Similarly, UVA8-based risk score can still separate the patients into low and high-risk groups in grade II or grade III gliomas (Fig. 4b). Although, IDH mutation status is a widely used prognostic and predictive biomarker, the UVA8-based risk score can also separate patients into two risk groups in patients presorted based on IDH mutation status (Fig. 4c). UVA8-derived risk score can also stratify patients into two risk groups among male and female patients (Fig. 4d).



Age	P-val	HR	CI
<40	4.65e-5	3.99	2.0-8.2
>=40	1.83e-9	7.20	3.4-15.2

Grade	P-val	HR	CI
II	2.8e-5	4.60	2.1-10.1
III	2.5e-7	5.21	2.6-10.5

IDH	P-val	HR	CI
WT	0.00025	3e+8	0-Inf
MUT	4.7e-6	3.36	1.9-5.8

Gender	P-val	HR	CI
Male	1.3e-8	5.50	2.8-10.6
Female	5.6e-7	5.75	2.7-12.4

Fig. 4 Stratification analysis by different clinical variables. Kaplan–Meier curve analysis of overall survival in high- and low-risk groups for **a** younger (age < 40) and older patients (age ≥ 40). **b** Grade II and grade III patients **c** IDH mutation status as WT and mutation (MUT) patients **d** male and female patients. Black-dashed line: patients with high-risk score, gray solid line: patients with low-risk score. The tables on the right show log-rank, *p* value, hazard ratio, and 95% confidence interval for each Kaplan–Meier plot

Conversely, we tested whether these standard clinically used parameters, age, gender, grade, and IDH mutation status, continue to independently stratify patients even after they have been presorted into two groups by UVA8 risk score (Fig. S7). In patients with high UVA8 risk score, age, grade, and IDH mutations status can further separate the patients into two groups of better or worse outcome. In contrast, in patients with low UVA8 risk scores, none of the clinical factors could further stratify patients into two different survival groups with a p value < 0.05 (Fig. S7). Consistent with the previous observation (Fig. S1), gender is ineffective in stratifying patients into two categories within patients with high- or low-risk score.

UVA8 is a Better Predictor of Glioma Patients' Survival

We assessed the accuracy of UVA8 in prediction of survival by comparing its time-dependent AUC in a ROC curve with that of other clinical characteristics. For each prognostic factor (e.g., UVA8, IDH status, etc.), we varied the cutoff so as to

vary the false positive rate for 5-year survival prediction from 0 to 1. For each cutoff, the corresponding true positive rate for 5-year survival was calculated (Fig. 5a). Comparing the AUC for these ROC curves suggested that UVA8 performs best in predicting survival of the glioma patients compared to the other criteria. This calculation was extended to predict survival of other durations (1–16 years) and the AUC plotted for each predictor (Fig. 5b). UVA8 can predict survival better for all durations, particularly at the very early years after diagnosis when the prediction is worse for most of the predictors. Since, gender is not associated with glioma patients' survival (Fig. S1), the prediction of outcome was no better than random guess (AUC = 0.5) (Fig. 5a, b). We employed Cox multivariable probability hazard model to identify the impact of UVA8 and different clinicopathological characteristics in estimating hazard (Fig. 5c). UVA8 is most significantly correlated with the survival information ($p = 1.4e\text{-}07$) and shows highest hazard ratio (HR = 4), indicating that the risk score performs better than any other currently used approaches for prognosis. Here, the hazard ratio of UVA8 is calculated by dichotomizing the risk score of > − 0.023 (median risk score from TCGA) to 1 and < − 0.023 to 0 to compare the hazard rates of high-risk versus low-risk patients. The hazard ratio of the eight lncRNAs individually and combined

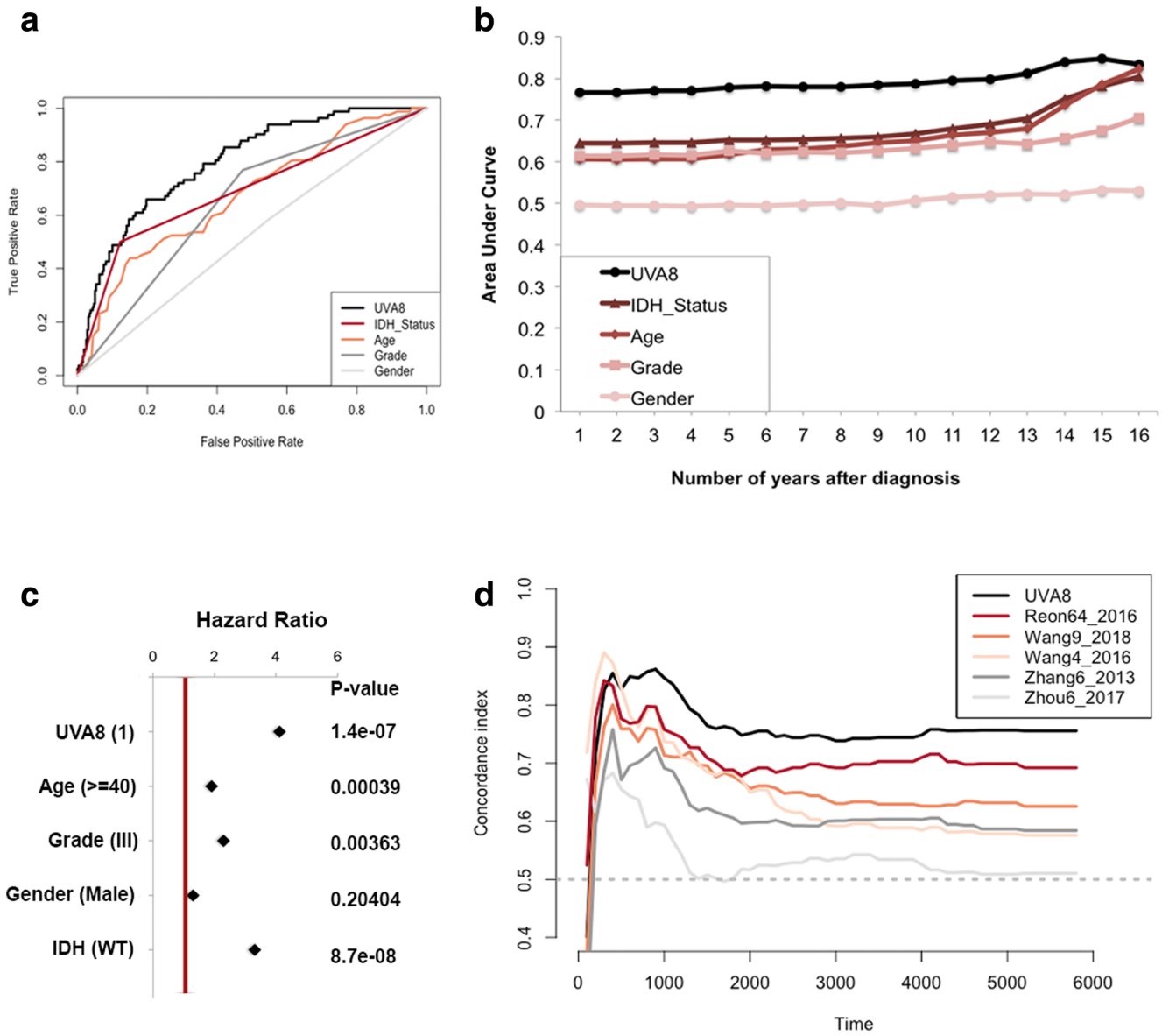


Fig. 5 Performance evaluation of the 8 lncRNA-based risk score. **a** Receiver operating characteristic curve for 5-year survival shows UVA8 has better area under curve compared with other predictors. **b** Area under curve plotted for different durations of survival for eight lncRNA-based risk score, tumor grade, age, IDH mutation status, and gender of patients in TCGA cohort. **c** Cox multivariate regression with clinical information and risk score calculated from UVA8 for survival in TCGA cohort. **d** Concordance index showing measure of concordance of predictor with survival of patients in TCGA

as risk score is tabulated in Supplementary Table S1. The UVA8 risk score is associated with more hazard (HR = 2) than any of the individual lncRNA supporting the importance of a combinatorial signature than an individual RNA for predicting survival. The hazard ratio of UVA8 in Supplementary Table S1 is different from that in Fig. 5c because in the former the hazard ratio is calculated with the risk score as a continuous variable.

We then sought to compare the performance of UVA8-based survival model with published lncRNA-based survival models by calculating Cindex (as discussed in "Materials and Methods") for TCGA dataset for each of the models. We first calculated risk score for each patient by considering the expression level of the prognostic lncRNAs in each model weighted by their estimated regression coefficients retrieved from the respective studies (Supplementary Table S2). The patients were ordered based on their actual survival at a given time after diagnosis and based on their risk score in each model. The concordance of the two orders is measured in pairwise comparisons of the patients to calculate a single time-dependent concordance index for the model that is being evaluated. This is repeated for different survival times with an interval of 100 days. The concordance index for each survival time for UVA8 and all published models is tabulated as

a

lncRNAs	Positively correlated (a,b,c)	Negatively correlated (a,b,c)
RP11-266K4.14	FGFBP3 (0.57,-0.39,3.6e-5)	CYSTM1 (-0.50,0.20,0.13)
FLJ37035	ZNF32 (0.60,-0.53,0.001)	CDC42 (-0.55,0.22.0.03)
LINC01561	PLPP4 (0.78,-0.23,0.018)	NR2E1 (-0.53,0.004,0.92)
RP11-118K6.3	RTP5 (0.61,-0.19,0.174)	DDOST (-0.62,0.23,0.027)
DGCR9	BRSK2 (0.74,-0.44,0.00017)	GNG5 (-0.66,0.45,8.5e-06)
RP11-142A22.3	PRRT1 (0.70,-0.15,0.25)	SMC5 (-0.68,0.22,0.006)
LINC00641	AKAP6 (0.68,0.07,0.51)	EDEM2 (-0.72,0.07,0.53)
RP11-96H19.1	NUDT7 (0.63,-0.311,0.021)	RTCA (-0.58,0.28,0.02)

a spearman correlation
b Cox-coefficient
c log-rank p-value

b

Hallmark Gene Sets	Normalized Enrichment Score
HALLMARK_INTERFERON_GAMMA_RESPONSE	4.76
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	4.38
HALLMARK_E2F_TARGETS	4.22
HALLMARK_INTERFERON_ALPHA_RESPONSE	3.79
HALLMARK_G2M_CHECKPOINT	3.74
HALLMARK_ALLOGRAFT_REJECTION	3.56
HALLMARK_TNFA_SIGNALING_VIA_NFKB	3.52
HALLMARK_INFLAMMATORY_RESPONSE	3.31
HALLMARK_IL6_JAK_STAT3_SIGNALLING	3.13
HALLMARK_COMPLEMENT	2.96

c

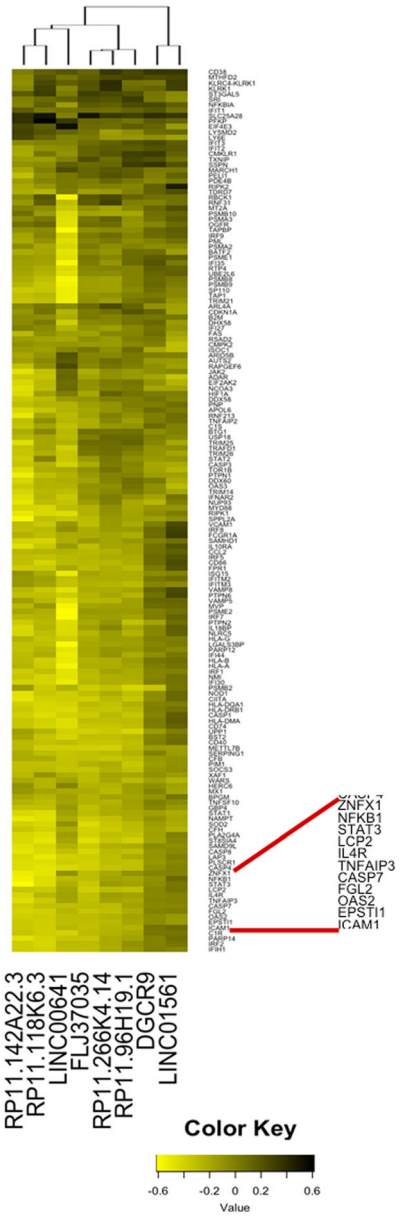


Fig. 6 Guilt-by-association analysis of the 8 lncRNAs in UVA8. **a** Correlation and Cox regression coefficient for the mRNAs that are most correlated (positive and negatively) with each of the lncRNAs in UVA8. a, b, and c defined below the table. **b** List of pathways that are most enriched in protein-coding genes that are negatively correlated with the UVA8 lncRNAs. **c** Heatmap showing correlation of different genes in the interferon gamma response gene set (rows) to the lncRNAs in UVA8 (columns)

Supplementary Table S3. UVA8 outperforms all existing lncRNA-based survival models at different times after diagnosis (Fig. 5d). As expected, prognostic signatures that were specific to GBMs (Zhang6_2013 and Zhou6_2017) show poor concordance index when used to predict survival of lower grade glioma patients.

Interferon Signaling is the Most Enriched Pathway in Guilt-by-Association with UVA8

Although many lncRNAs have been identified there has been very little functional annotation of the RNAs. We therefore applied guilt-by-association to infer functions of the lncRNAs associated with survival in UVA8. First, we interrogated whether protein-coding genes most correlated with an lncRNA in TCGA glioma cohort are themselves predictive of outcome. All the lncRNAs in UVA8 are associated with a negative Cox coefficient (protective). Of the eight mRNAs most correlated positively with these eight lncRNAs, five also have a negative Cox coefficient with a significant p value. Conversely, of the eight mRNAs most anti-correlated with these lncRNAs, five have a positive Cox coefficient with a significant p value (Fig. 6a). This result is consistent with the expectation that the expression of these protective lncRNAs will be positively correlated with expression of protective mRNAs and negatively correlated with the expression of harmful mRNAs.

GSEA analysis on protein-coding genes pre-ranked from most positively correlated to most negatively correlated to the lncRNA revealed several common pathways co-regulated with each of the eight lncRNAs (Fig. 6b). Interestingly, among the mRNAs that are negatively correlated with the lncRNAs, genes involved in immune and inflammatory response (IFNG, IFNA, allograft rejection, NFκB inflammatory response, and JAK-STAT pathway) are highly enriched. Similarly, genes involved in epithelial to mesenchymal transition and cell-cycle progressions are also most enriched. These gene set enrichments suggest a conventional tumor suppressor phenotype associated with these eight lncRNAs.

Many of the mRNAs are common in the IFNG, IFNA, allograft rejection, NFκB inflammatory response, and JAK-STAT gene sets. The genes upregulated in response to IFNG are mostly negatively correlated to lncRNAs in UVA8. To visualize this, the correlation coefficients were plotted for each lncRNA (columns) with individual mRNAs in the IFNG response pathway (rows) (Fig. 6c). Out of eight, six lncRNAs (RP11-266K4.14, FLJ37035, RP11-118K6.3, RP11-142A22.3, LINC00641, and RP11-96H19.1) are clustered together because they are more negatively correlated with genes of interferon gamma response pathway (Fig. 6c).

We found both NFκB and STAT3 genes as highly negatively correlated with the expression of the protective lncRNAs in UVA8. Genes involved in epithelial to mesenchymal transition and encoding cell cycle-related targets of E2F transcription factors and involved in G2/M checkpoints were also negatively correlated with UVA8 expression. On the other hand, genes that are downregulated upon activation of the oncogenes KRAS are positively correlated with the expression of the protective lncRNAs of UVA8.

eRNA (enhancer RNA) are another class of long non-coding RNAs which are 50–2000 bases long, unspliced, and non-polyA non-coding RNA expressed from enhancers involved in the activation of distantly located genes [41]. In order to check whether these lncRNAs can possibly act as eRNAs, we also checked the distance between lncRNAs and their correlated genes and found that these lncRNAs are correlated to several genes located in different location of genome suggesting a trans-regulation by these lncRNAs (data not shown). More experimental studies are required in future to decipher the role of these lncRNAs in regulating these genes and whether this regulation explains the effect of the lncRNAs on glioma tumor progression.

In order to investigate whether somatic mutations in these lncRNAs might account for the prognostic ability of the model, we used the somatic variant calls from The Cancer Genome Atlas ("Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines"). Based on this dataset, there seems to be somatic mutations in these lncRNAs in other cancers (most prominently DGCR9), but no somatic mutations were found in the eight lncRNAs in the TCGA lower grade glioma patients. Therefore, we feel that the predictive ability of our model is due to differences in the expression of these lncRNAs and not acquired somatic mutations in these lncRNAs. To see if copy number variation was the mechanism driving the differences in the expression of these eight lncRNAs, we tested whether the copy number of each gene was predictive of survival. We then correlated the expression of the gene to the copy number. Thus, copy number variation by itself may be predictive for two of the lncRNAs (FLJ37035 and LINC01561), but oddly for the second, there is no correlation between the CNV and level of expression. Thus, CNV is not the explanation for the expression differences of the lncRNAs, and is not a better predictor for prognosis (Supplementary Table S4).

Discussion

Gene expression profile reflects the underlying biological processes of disease. Cox regression is a widely used

approach to decipher correlation between gene expression profile and patient outcome. Previous analyses on microarray data explored protein-coding genes that could predict the prognosis of gliomas, particularly focusing on high-grade GBMs. LncRNAs are a class of RNA which can serve as a better prognostic marker than protein-coding mRNAs because they are numerous and cell-type specific [2, 3]. Additionally, since lncRNAs do not encode protein, they are the ultimate effectors, and their expression levels more accurately predict the levels of their activity. Recent studies have detected tumor-specific lncRNAs in exosomes, apoptotic bodies, and microparticles highlighting another advantage of considering lncRNAs in tumors, because they are expected to appear as fluid-based markers for the diagnosis of different cancers [42–44]. Among six published lncRNA-based prognostic signatures for gliomas, two are for predicting outcome in GBMs and one specifically for anaplastic gliomas. Wang et al. and Chen et al. have shown that a set of only four lncRNAs could predict survival in gliomas [23, 25]. However, the sequence of one of the lncRNAs in Chen et al., CR613436, was removed by the submitter on NCBI. Recently, the role of immune-related genes in glioma malignancies is gaining attention leading to the discovery of immune-related lncRNA-based prognostic markers for GBMs and anaplastic gliomas [22, 45]. Remarkably, there is no overlap between the prognostic lncRNAs identified in the aforementioned studies. Moreover, these studies are based on microarray data raising concerns particular to hybridization-based approaches, including reliance on current knowledge of expressed genes, problems of cross-hybridization, and cross-experiment comparison. Another issue is that association of lncRNAs with survival using Cox regression was sometimes carried out without controlling for any dependent variables and without penalizing for the effect of large number of variables.

In the present study, we have used an approach to screen lncRNAs from high-dimensional TCGA RNA-seq data, which is one of the largest and the most updated data for lower grade gliomas. After controlling for effects like age, grade, gender, and IDH mutation status, we applied regularization to penalize the effect of many dependent variables and select the lncRNAs based on 100 trials. We showed the robustness of eight lncRNA-based predictors in a completely independent cohort of Chinese glioma patients. The lncRNA prognostic signature identified in the present study, UVA8, is an independent predictor of survival in TCGA glioma patients. Since UVA8 is also a better predictor than the few patient and molecular characteristics currently used for prognosis in the clinic, a simple RNA quantification will aid the physician to decide whether to adopt more aggressive therapy at the outset.

The protective lncRNAs that constitute UVA8 are negatively correlated with protein-coding genes involved in interferon gamma and inflammatory response highlighting the role of immune-response genes in glioma progression. Except LINC01561, all seven lncRNAs (RP11-266K4.14, FLJ37035, RP11-118K6.3, DGCR9, RP11-142A22.3, LINC00641, and RP11-96H19.1) are negatively correlated to most of the protein-coding genes, which are upregulated in response to interferon gamma/alpha, genes regulated by NFκB in response to TNF, inflammatory response, and genes upregulated by IL6 via STAT3. This suggests that an active immune reaction perhaps in response to cytokines secreted from tumor and immune cells is predictive of poor outcome in gliomas. NFκB and JAK/STAT pathways are known to be aberrantly upregulated in GBMs. The level of NFκB increases as the tumors progress in astrocytic tumors [46, 47] and STAT3 is constitutively active in GBMs [48, 49]. Immune-related pathways are also known to be involved in glioma tumor cell proliferation [50], survival [45], invasion [51], and chemoresistance [52]. In addition, epithelial-mesenchymal transition (associated with invasion) and active cell proliferation are suppressed if UVA8 lncRNAs are high, and this leads to better outcome, consistent with our understanding of how invasion and cell proliferation negatively impact outcome. On the other hand, genes that were positively correlated with the expression of UVA8 are enriched in genes that are down regulated by activation of the oncogene KRAS.

There are reports of the same lncRNA being predictive of outcome in the same manner in multiple tumor types. For example, DRAIC expression predicts good outcome in gliomas, melanomas, and cancers of the prostate, stomach, liver, kidney, and lung [53]. In contrast, expression of LINC00152/CYTOR is predictive of poor outcome in gliomas, and cancers of the head and neck, lung, kidney, liver, and pancreas (our unpublished work). Such observations are particularly exciting because they imply that the lncRNA has an important role in tumor biology that transcends tumor types, and these RNAs should be prioritized for cell- and molecular-biology studies to discern their function. It will thus be very interesting to explore whether any of the lncRNAs of UVA8 will be protective in other tumor types. Finally, future studies will address whether structural variation, copy number variations, and sequence polymorphism of these lncRNAs contribute to the prognostic outcome. We are excited that UVA8 was also predictive of outcome in a completely different tumor cohort (CGGA) from a patient population that is from an entirely different geographical location with attendant differences in environment and population genotypes. It will be interesting to see if UVA8 is equally predictive of outcome in other patient populations from other parts of the world.

Acknowledgments We thank Dr. Stefan Bekiranov, Dr. William Pearson, and Dutta lab members for helpful discussions. M.K. is supported by a DOD award PC151085.

Funding Information The work was supported by a V foundation award D2018-002 and R01 AR067712 from NIAMS.

References

1. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22:1775–1789. https://doi.org/10.1101/gr.132159.111
2. Cabili M, Trapnell C, Goff L et al (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25:1915–1927. https://doi.org/10.1101/gad.17446611
3. Huarte M (2015) The emerging role of lncRNAs in cancer. Nat Med 21:1253–1261. https://doi.org/10.1038/nm.3981
4. Schmitt AM, Chang HY (2016) Long noncoding RNAs in cancer pathways. Cancer Cell 29:452–463. https://doi.org/10.1016/j.ccell.2016.03.010
5. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA et al (2005) Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 352:987–996. https://doi.org/10.1056/NEJMoa043330
6. Huang J, Samson P, Perkins SM, Ansstas G, Chheda MG, DeWees TA, Tsien CI, Robinson CG et al (2017) Impact of concurrent chemotherapy with radiation therapy for elderly patients with newly diagnosed glioblastoma: a review of the National Cancer Data Base. J Neuro-Oncol 131:593–601. https://doi.org/10.1007/s11060-016-2331-6
7. Ducray F, Idbaih A, Wang X-W, Cheneau C, Labussiere M, Sanson M (2011) Predictive and prognostic factors for gliomas. Expert Rev Anticancer Ther 11:781–789. https://doi.org/10.1586/era.10.202
8. Carninci P, Kasukawa T, Katayama S et al (2005) The transcriptional landscape of the mammalian genome. Science 309:1559–1563. https://doi.org/10.1126/science.1112014
9. Ravasi T, Suzuki H, Pang KC, Katayama S, Furuno M, Okunishi R, Fukuda S, Ru K et al (2006) Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res 16:11–19. https://doi.org/10.1101/gr.4200206
10. Mehler MF, Mattick JS (2007) Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. Physiol Rev 87:799–823. https://doi.org/10.1152/physrev.00036.2006
11. Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS (2010) Non-coding RNAs: regulators of disease. J Pathol 220:126–139
12. Qureshi IA, Mattick JS, Mehler MF (2010) Long non-coding RNAs in nervous system function and disease. Brain Res 1338:20–35
13. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS (2008) Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci U S A 105:712–716. https://doi.org/10.1073/pnas.0706729105
14. Amaral PP, Neyt C, Wilkins SJ, Askarian-Amiri ME, Sunkin SM, Perkins AC, Mattick JS (2009) Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. RNA 15:2013–2027. https://doi.org/10.1261/rna.1705309
15. Johnson R, Teh CH-L, Jia H, Vanisri RR, Pandey T, Lu ZH, Buckley NJ, Stanton LW et al (2009) Regulation of neural macroRNAs by the transcriptional repressor REST. RNA 15:85–96. https://doi.org/10.1261/rna.1127009
16. Arron JR, Winslow MM, Polleri A, Chang CP, Wu H, Gao X, Neilson JR, Chen L et al (2006) NFAT dysregulation by increased dosage of DSCR1 and DYRK1A on chromosome 21. Nature 441:595–600. https://doi.org/10.1038/nature04678
17. Wang J, Zhao H, Fan Z, Li G, Ma Q, Tao Z, Wang R, Feng J et al (2017) Long noncoding RNA H19 promotes neuroinflammation in ischemic stroke by driving histone deacetylase 1-dependent M1 microglial polarization. Stroke 48:2211–2221. https://doi.org/10.1161/STROKEAHA.117.017387
18. Chubb JE, Bradshaw NJ, Soares DC, Porteous DJ, Millar JK (2008) The DISC locus in psychiatric illness. Mol Psychiatry 13:36–64. https://doi.org/10.1038/sj.mp.4002106
19. Zhang X, Sun S, Pu JKS, Tsang ACO, Lee D, Man VOY, Lui WM, Wong STS et al (2012) Long non-coding RNA expression profiles predict clinical phenotypes in glioma. Neurobiol Dis 48:1–8. https://doi.org/10.1016/J.NBD.2012.06.004
20. Reon BJ, Anaya J, Zhang Y, Mandell J, Purow B, Abounader R, Dutta A (2016) Expression of lncRNAs in low-grade gliomas and glioblastoma multiforme: an in silico analysis. PLoS Med 13:e1002192. https://doi.org/10.1371/journal.pmed.1002192
21. Li R, Qian J, Wang Y-Y, Zhang JX, You YP (2014) Long noncoding RNA profiles reveal three molecular subtypes in glioma. CNS Neurosci Ther 20:339–343. https://doi.org/10.1111/cns.12220
22. Wang W, Zhao Z, Yang F, Wang H, Wu F, Liang T, Yan X, Li J et al (2018) An immune-related lncRNA signature for patients with anaplastic gliomas. J Neuro-Oncol 136:263–271. https://doi.org/10.1007/s11060-017-2667-6
23. Wang W, Yang F, Zhang L et al (2016) LncRNA profile study reveals four-lncRNA signature associated with the prognosis of patients with anaplastic gliomas. Oncotarget 7:77225–77236. https://doi.org/10.18632/oncotarget.12624
24. Zhang X-Q, Sun S, Lam K-F, Kiang KMY, Pu JKS, Ho ASW, Lui WM, Fung CF et al (2013) A long non-coding RNA signature in glioblastoma multiforme predicts survival. Neurobiol Dis 58:123–131. https://doi.org/10.1016/J.NBD.2013.05.011
25. Chen G, Cao Y, Zhang L et al (2017) Analysis of long non-coding RNA expression profiles identifies novel lncRNA biomarkers in the tumorigenesis and malignant progression of gliomas. Oncotarget 8:67744–67753. https://doi.org/10.18632/oncotarget.18832
26. van de Vijver MJ, He YD, van't Veer LJ et al (2002) A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347:1999–2009. https://doi.org/10.1056/NEJMoa021967
27. Spentzos D, Levine D, Ramoni M et al (2004) Gene expression signature with independent prognostic significance in epithelial ovarian cancer. J Clin Oncol 22:4700–4710. https://doi.org/10.1200/jco.2004.04.070
28. Bullinger L, Döhner K, Bair E, Fröhling S, Schlenk RF, Tibshirani R, Döhner H, Pollack JR (2004) Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 350:1605–1616. https://doi.org/10.1056/NEJMoa031046
29. Chibon F (2013) Cancer gene expression signatures-the rise and fall? Eur J Cancer 49:2000–2009. https://doi.org/10.1016/j.ejca.2013.02.021
30. Bao ZS, Chen HM, Yang MY, Zhang CB, Yu K, Ye WL, Hu BQ, Yan W et al (2014) RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. Genome Res 24:1765–1773. https://doi.org/10.1101/gr.165126.113
31. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D et al (2012) GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 22:1760–1774. https://doi.org/10.1101/gr.135350.111

32. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33:290–295. https://doi.org/10.1038/nbt.3122
33. Tibshirani R (1997) The lasso method for variable selection in the cox model. Stat Med 16:385–395. https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3
34. Tibshirani R (2011) Regression shrinkage and selection via the lasso: a retrospective. J R Stat Soc Ser B Stat Methodol 73:273–282. https://doi.org/10.1111/j.1467-9868.2011.00771.x
35. Goeman JJ (2010) L1 penalized estimation in the Cox proportional hazards model. Biom J 52:70–84. https://doi.org/10.1002/bimj.200900028
36. Alizadeh AA, Gentles AJ, Alencar AJ, Liu CL, Kohrt HE, Houot R, Goldstein MJ, Zhao S et al (2011) Prediction of survival in diffuse large B-cell lymphoma based on the expression of 2 genes reflecting tumor and microenvironment. Blood 118:1350–1358. https://doi.org/10.1182/blood-2011-03-345272
37. Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D, Levy R (2004) Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 350:1828–1837. https://doi.org/10.1056/NEJMoa032520
38. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33: . doi: https://doi.org/10.18637/jss.v033.i01
39. Raykar VC, Steck H, Krishnapuram B, et al On ranking in survival analysis: bounds on the concordance index
40. Gerds TA, Kattan MW, Schumacher M, Yu C (2013) Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. Stat Med 32:2173–2184. https://doi.org/10.1002/sim.5681
41. Kim TK, Hemberg M, Gray JM (2015) Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. Cold Spring Harb Perspect Biol 7:a018622. https://doi.org/10.1101/cshperspect.a018622
42. Panzitt K, Tschernatsch MMO, Guelly C, Moustafa T, Stradner M, Strohmaier HM, Buck CR, Denk H et al (2007) Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. Gastroenterology 132:330–342. https://doi.org/10.1053/J.GASTRO.2006.08.026
43. Du Z, Fei T, Verhaak RGW et al (2013) Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. Nat Struct Mol Biol 20:908–913. https://doi.org/10.1038/nsmb.2591
44. Mohankumar S, Patel T (2016) Extracellular vesicle long noncoding RNA as potential biomarkers of liver cancer. Brief Funct Genomics 15:249–256. https://doi.org/10.1093/bfgp/elv058
45. Zhou M, Zhang Z, Zhao H, et al (2017) An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. Mol Neurobiol 1–14
46. Angileri FF, Aguennouz M, Conti A et al (2008) Nuclear factor-κB activation and differential expression of survivin and Bcl-2 in human grade 2-4 astrocytomas. Cancer 112:2258–2266. https://doi.org/10.1002/cncr.23407
47. Korkolopoulou P, Levidou G, Saetta AA, el-Habr E, Eftichiadis C, Demenagas P, Thymara I, Xiromeritis K et al (2008) Expression of nuclear factor-κB in human astrocytomas: relation to pIκBa, vascular endothelial growth factor, Cox-2, microvascular characteristics, and survival. Hum Pathol 39:1143–1152. https://doi.org/10.1016/J.HUMPATH.2008.01.020
48. Schaefer LK, Ren Z, Fuller GN, Schaefer TS (2002) Constitutive activation of Stat3α in brain tumors: localization to tumor endothelial cells and activation by the endothelial tyrosine kinase receptor (VEGFR-2). Oncogene 21:2058–2065. https://doi.org/10.1038/sj.onc.1205263
49. Abou-Ghazal M, Yang DS, Qiao W, Reina-Ortiz C, Wei J, Kong LY, Fuller GN, Hiraoka N et al (2008) The incidence, correlation with tumor-infiltrating inflammation, and prognosis of phosphorylated STAT3 expression in human gliomas. Clin Cancer Res 14:8228–8235. https://doi.org/10.1158/1078-0432.CCR-08-1329
50. Puliyappadamba VT, Hatanpaa KJ, Chakraborty S, Habib AA (2014) The role of NF-κB in the pathogenesis of glioma. Mol Cell Oncol 1:e963478. https://doi.org/10.4161/23723548.2014.963478
51. Kesanakurti D, Chetty C, Rajasekhar Maddirela D, Gujrati M, Rao JS (2013) Essential role of cooperative NF-κB and Stat3 recruitment to ICAM-1 intronic consensus elements in the regulation of radiation-induced invasion and migration in glioma. Oncogene 32:5144–5155. https://doi.org/10.1038/onc.2012.546
52. Coupienne I, Bontems S, Dewaele M, Rubio N, Habraken Y, Fulda S, Agostinis P, Piette J (2011) NF-kappaB inhibition improves the sensitivity of human glioblastoma cells to 5-aminolevulinic acid-based photodynamic therapy. Biochem Pharmacol 81:606–616. https://doi.org/10.1016/J.BCP.2010.12.015
53. Sakurai K, Reon BJ, Anaya J, Dutta A (2015) The lncRNA DRAIC/PCAT29 locus constitutes a tumor-suppressive nexus. Mol Cancer Res 13:828–838. https://doi.org/10.1158/1541-7786.MCR-15-0016-T

AMERICAN SOCIETY FOR MICROBIOLOGY | Molecular and Cellular Biology®

RESEARCH ARTICLE

MUNC, an Enhancer RNA Upstream from the *MYOD* Gene, Induces a Subgroup of Myogenic Transcripts in *trans* Independently of MyoD

Magdalena A. Cichewicz,[a] Manjari Kiran,[a] Róża K. Przanowska,[a] Ewelina Sobierajska,[a] Yoshiyuki Shibata,[a] Anindya Dutta[a]

[a]Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, Virginia, USA

ABSTRACT MyoD upstream noncoding RNA (MUNC) initiates in the distal regulatory region (DRR) enhancer of *MYOD* and is formally classified as an enhancer RNA (DRR^{eRNA}). MUNC is required for optimal myogenic differentiation, induces specific myogenic transcripts in *trans* (*MYOD*, *MYOGENIN*, and *MYH3*), and has a functional human homolog. The vast majority of eRNAs are believed to act in *cis* primarily on their neighboring genes (1, 2), making it likely that MUNC action is dependent on the induction of *MYOD* RNA. Surprisingly, MUNC overexpression in *MYOD*$^{-/-}$ C2C12 cells induces many myogenic transcripts in the complete absence of MyoD protein. Genomewide analysis showed that, while many genes are regulated by MUNC in a MyoD-dependent manner, there is a set of genes that are regulated by MUNC, both upward and downward, independently of MyoD. MUNC and MyoD even appear to act antagonistically on certain transcripts. Deletion mutagenesis showed that there are at least two independent functional sites on the MUNC long noncoding RNA (lncRNA), with exon 1 more active than exon 2 and with very little activity from the intron. Thus, although MUNC is an eRNA of *MYOD*, it is also a *trans*-acting lncRNA whose sequence, structure, and cooperating factors, which include but are not limited to MyoD, determine the regulation of many myogenic genes.

KEYWORDS MUNC, MyoD, eRNA, enhancer, lncRNA, myogenesis, skeletal muscles

Myogenesis is a process of skeletal muscle differentiation occurring during vertebrate embryo development and during regeneration of muscle fibers after injury in the adult. During embryonic development, muscles derive from the mesoderm, where myoblasts, embryonic progenitor cells, give rise to muscle fibers (3). Myogenesis requires a network of muscle-specific transcription factors composed of four muscle regulatory factors (MRFs) from the basic helix-loop-helix (bHLH) family of transcription factors (myogenic factor 5 [Myf5], myoblast determination protein [MyoD], myogenin, and muscle-specific regulatory factor 4 [MRF4]). When myogenesis is activated, MyoD-MyoE protein heterodimers bind to E-box sequences in promoters of genes, driving their transcription and setting off a transcriptional cascade (4). This activation leads to the expression of several muscle-specific target genes, such as *MYOGENIN*, *M-CADHERIN*, myosin heavy and light chains (such as *MYH3*), and the muscle creatine kinase gene (5).

Three DNA sequence elements regulate *MYOD* expression in mice: a proximal regulatory region (PRR) that is adjacent to the transcription start site (TSS) of *MYOD*, a 720-bp-long distal regulatory region (DRR) located ~5 kb upstream from the *MYOD* TSS, and a core enhancer region (CER) located ~23 kb upstream from the *MYOD* TSS (6–8). The DRR sequence is functionally conserved between mouse and human, sharing blocks of sequence identity over a 445-bp region between the two species. DRR deletion reduces *MYOD* RNA and the protein level in adult muscle (9, 10). The DRR

Received 17 December 2017 **Returned for modification** 2 January 2018 **Accepted** 13 July 2018

Accepted manuscript posted online 23 July 2018

Citation Cichewicz MA, Kiran M, Przanowska RK, Sobierajska E, Shibata Y, Dutta A. 2018. MUNC, an enhancer RNA upstream from the *MYOD* gene, induces a subgroup of myogenic transcripts in *trans* independently of MyoD. Mol Cell Biol 38:e00655-17. https://doi.org/10.1128/MCB.00655-17.

Address correspondence to Anindya Dutta, ad8q@virginia.edu.

M.A.C., M.K., and R.K.P. contributed equally to this study.

contains consensus binding sites for MyoD, MEF-2, and SRF (10, 11), explaining how it positively regulates *MYOD* expression like a classic enhancer. The DRR is essential as an enhancer for skeletal muscle differentiation, but it also serves as the initiation site of a myogenic enhancer RNA (eRNA), MyoD upstream noncoding RNA (MUNC), or DRR^{eRNA}, which plays a positive regulatory role during muscle development (12, 13).

Long noncoding RNAs (lncRNAs) form a diverse family of RNA transcripts longer than 200 nucleotides (nt) that do not encode proteins but have different functions in the cell as RNA molecules (reviewed in reference 14). High-throughput RNA sequencing (RNA-Seq) analysis in mice suggests that lncRNAs are a major component of the transcriptome (15). Mainly transcribed by RNA polymerase II (RNA Pol II), lncRNA can be intergenic, multiexonic, antisense to known genes, or from regulatory elements located distal to a known TSS. High-throughput RNA sequencing identified many novel lncRNAs specifically expressed during skeletal muscle differentiation (16). Their mechanisms of action are heterogeneous, and they are localized differently in cells (reviewed in references 14 and 17). Nuclear lncRNAs can mediate epigenetic changes by recruiting chromatin-remodeling complexes to specific genomic loci. Muscle-specific steroid receptor RNA activator (SRA) RNA promotes muscle differentiation through its interactions with RNA helicase coregulators p68, p72, and MyoD (18). Another example of a promyogenic lncRNA functioning in *cis* is Dum (developmental pluripotency-associated 2 [Dppa2] upstream binding muscle RNA), which silences its neighboring gene, *DPPA2*, by recruiting Dnmts to its locus (19). DBE-T, a lncRNA produced selectively in patients with facioscapulohumeral muscular dystrophy (FSHD), binds to the chromatin and recruits transcriptional activator Ash1L to derepress the *FSHD* locus (20).

An important group of nuclear lncRNAs work as eRNAs, stimulating transcription of adjacent genes (1). A recent study of 12 mouse lncRNAs identified 5 of them that act as eRNAs stimulating the transcription of the adjoining gene in *cis* by a process that involves the transcription and splicing of the eRNA but is not dependent on the sequence of the actual RNA transcript (2). Myogenic eRNAs include DRR^{eRNA}, or MUNC, and CER^{eRNA}, which, consistent with current models of eRNA function, stimulate expression of the adjoining *MYOD* gene in *cis* by increasing chromatin accessibility for transcriptional factors. DRR^{eRNA}, or MUNC, is already a little atypical as an eRNA because it can induce expression not only of the *MYOD* gene located in *cis* but also of *MYOGENIN* and *MYH3*, which are located on different chromosomes (12, 13).

In this study, we show that MUNC has a function independent of its action as an eRNA stimulating expression of *MYOD*. Specifically, MUNC has a MyoD-independent promyogenic function during skeletal muscle differentiation, has multiple separate functional regions, and can act in *trans* on multiple genes on different chromosomes. These findings raise the possibility that, although many eRNAs act as classic enhancer RNAs that stimulate transcription of adjoining genes merely by the acts of transcription and splicing, some of them have additional roles as *trans*-acting lncRNAs, where the sequence of the RNA matters for its function.

RESULTS

MUNC as a lncRNA has multiple domains important for its function. In the previous study, we showed that stable overexpression of MUNC from a heterologous site in C2C12 cells increases the levels of three myogenic RNAs, *MYOD*, *MYOGENIN*, and *MYH3* (13). This in itself is at odds with the prevailing model, in which the acts of transcription and splicing at the endogenous eRNA locus are important for the action of the eRNA. We therefore decided to investigate the second tenet of the eRNA hypothesis: is the specific sequence of the MUNC transcript irrelevant for stimulating the myogenic transcripts? Fragments of MUNC containing different parts of the RNA were stably overexpressed in C2C12 cells (Fig. 1A). The overexpression was confirmed both in proliferating myoblasts (Fig. 1C to E) and in differentiating myotubes (Fig. 1F to H). In addition, we used C2C12 cells stably transfected with the spliced isoform of MUNC and with the genomic sequence of MUNC (overexpressing both spliced and unspliced isoforms). We compared the expression levels of *MYOD*, *MYOGENIN*, and

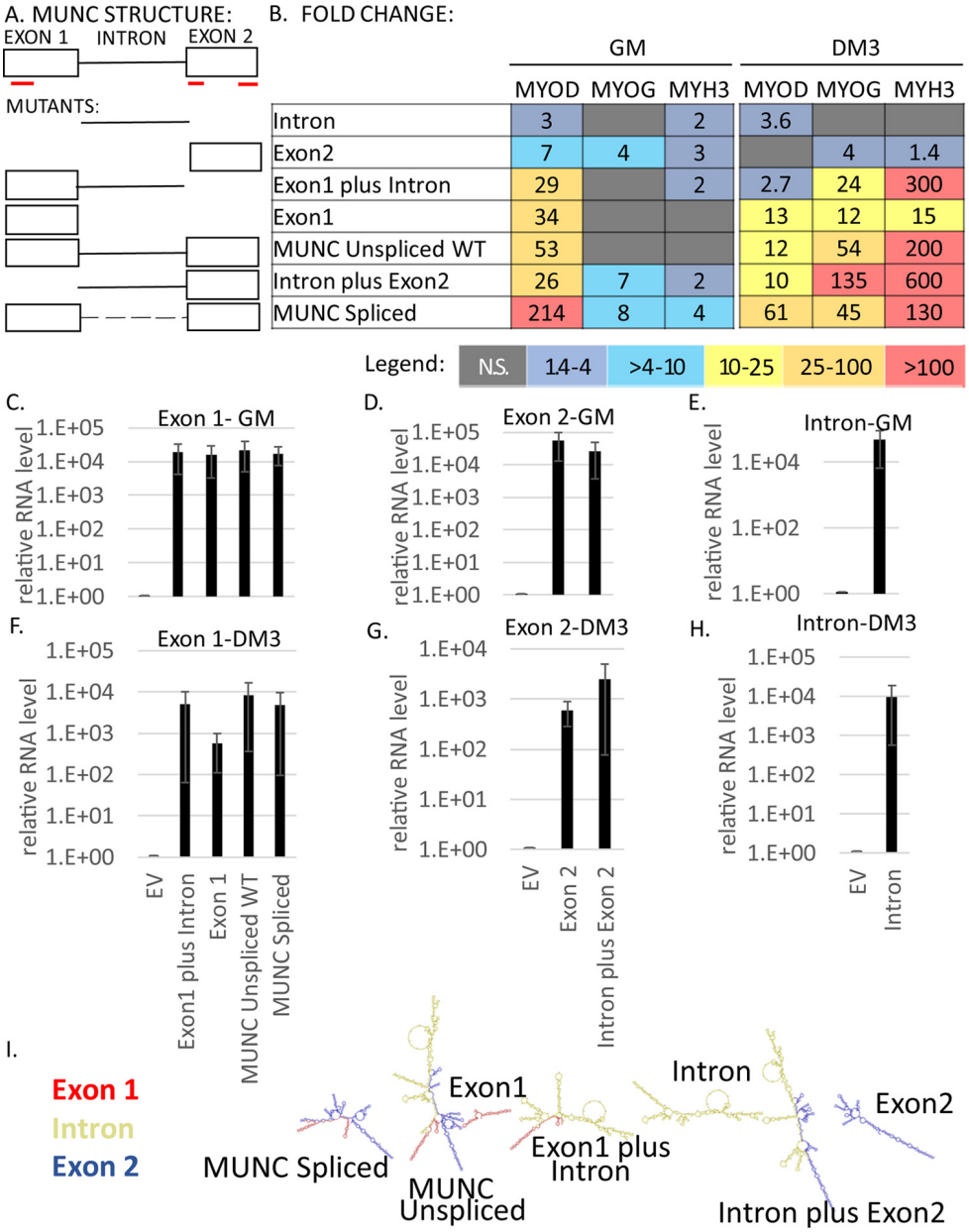


FIG 1 MUNC has at least two domains important for its function. (A) Schematic illustrating MUNC structure. The red lines indicate three potential micropeptides coded by MUNC spliced sequence: two of 20 amino acids and one of 60 amino acids. The micropeptides were defined using a translation tool (http://web.expasy.org/translate/). (B) Heat maps showing summaries of qRT-PCR analyses of C2C12 mutant cells stably overexpressing different truncated MUNC sequences. Levels of myogenic factor transcripts were measured in three biological runs and normalized to the GAPDH (glyceraldehyde-3-phosphate dehydrogenase) level and to control cells under each condition, and mean values were calculated. The colors used in the heat maps correspond to fold changes according to the legend. N.S., not significant. Analysis of proliferating cells and differentiating cells. (C to H) qRT-PCR analysis of mutant cells overexpressing truncated MUNC sequences showing levels of different parts of the transcript (exon 1, intron, and exon 2) in GM (C to E) and in DM3 (F to H). The data were normalized to GAPDH and to control cells transfected with an EV. The values represent three biological replicates and are presented as means and standard errors of the mean (SEM). (I) Predicted structures of different mutants of MUNC generated using the Forna RNA prediction tool.

MYH3 RNAs in cells overexpressing MUNC or fragments of MUNC relative to control cells transfected with the empty vector (EV). We performed the analysis under two conditions: in proliferating myoblasts (growth medium [GM]) to see whether MUNC is able to induce myogenic factors when cells proliferate, and after 3 days of differentiation (DM3) in differentiation medium (DM) to see whether overexpression of MUNC is

still able to change myogenic RNA levels when other myogenic factors have already been induced (Fig. 1B). Several interesting points emerge from consideration of the results.

First, in differentiating cells, MUNC induced *MYOGENIN* and *MYH3* to much higher levels than in proliferating cells, suggesting that differentiating cells may express additional factors that facilitate MUNC's action. Second, *MYOD* induction by exon 1, intron plus exon 2, or unspliced or spliced MUNC was much lower in DM3 (10 to 61 times than in cells without MUNC overexpression) than in GM (26 to 214 times), yet the reverse was true for *MYOGENIN* and *MYH3* (12 to 600 times in DM3 versus 1 to 8 times in GM). This suggests that there is not a linear correlation between the fold induction of *MYOD* and that of *MYOGENIN* and *MYH3*, as would have been expected if MUNC worked solely by inducing *MYOD* to induce *MYOGENIN* or *MYH3*. This lack of correlation is consistent with our earlier observation that MUNC overexpression induced *MYOGENIN* and *MYH3* mRNAs without inducing MyoD protein (despite the induction of *MYOD* mRNA) (13).

Third, spliced MUNC was always better than genomic MUNC at inducing *MYOD*. We know from RNA-Seq that genomic MUNC expresses mostly unspliced MUNC in these cells, so the difference is probably attributable either to the presence of inhibitory sequences in the intron or to different folding of the exonic sequences in unspliced and spliced MUNC. Differences in folding of the two isoforms were predicted by the Forna tool (21) (Fig. 1I). Among truncated mutants of MUNC, exon 1 was the most potent at inducing *MYOD*. Although the intron and exon 2 by themselves were mostly ineffective, addition of the intron to exon 2 made it more effective at inducing *MYOD* than either of them alone. As Fig. 1I shows, exon 2, intron, and exon 2 plus intron fragments of MUNC have different predicted RNA-folding structures.

In summary, these studies suggest that the simple act of transcription of MUNC (as suggested for eRNAs) cannot be enough for the stimulation of *MYOD*, *MYOGENIN*, or *MYH3*. Instead, as isolated fragments, exon 1 has the most significant stimulatory activity, although a second domain with activity became evident in the intron plus exon 2 fragment. Finally, the high degree of activity of the intron plus exon 2 fragment compared to either part alone (intron or exon 2) or of spliced MUNC (exon 1 plus exon 2) compared to unspliced MUNC (exon 1 plus intron plus exon 2) suggests that the folding of the RNA is important for this activity.

There have been a few reports of lncRNAs encoding micropeptides with biological functions (22). Spliced MUNC transcripts could code for three such micropeptides unrelated to each other in sequence (underlined in red in Fig. 1A). The structure-function analysis mentioned above rules out the possibility that the induction of the three genes is due to any of these micropeptides.

MYOD knockout (KO) diminishes muscle differentiation *in vitro*. A crucial role of MyoD during skeletal muscle differentiation was established both *in vitro* and *in vivo*. Skeletal muscles of *MYOD*$^{-/-}$ mice displayed reduced capacity for regeneration following injury (23), and *in vitro* knockdown of *MYOD* in differentiating C2C12 cells decreased the efficiency of differentiation (13, 24). It is also known that knockdown of MUNC decreases expression of *MYOD* and negatively affects other downstream effectors of muscle differentiation (13). To investigate whether the role of MUNC during muscle differentiation is through the induction of MyoD or whether MUNC has activities independent of MyoD, we engineered *MYOD*$^{-/-}$ C2C12 cells. Using clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 technology (25), both alleles of *MYOD* were knocked out by deletion of 149 bp of *MYOD* exon 1 (corresponding to amino acids P7 to L57 of the MyoD protein and throwing the rest of the protein out of frame) (Fig. 2A). The deletion was confirmed by PCR of the genomic DNA (Fig. 2B) and by Sanger sequencing of the PCR products (Fig. 2A). RNA-Seq data provided additional corroboration of the deletion by showing the complete absence of reads from the deleted region in *MYOD*$^{-/-}$ cells compared to wild-type (WT) cells (Fig. 2C). The homozygous deletion was associated with the complete absence of MyoD protein in

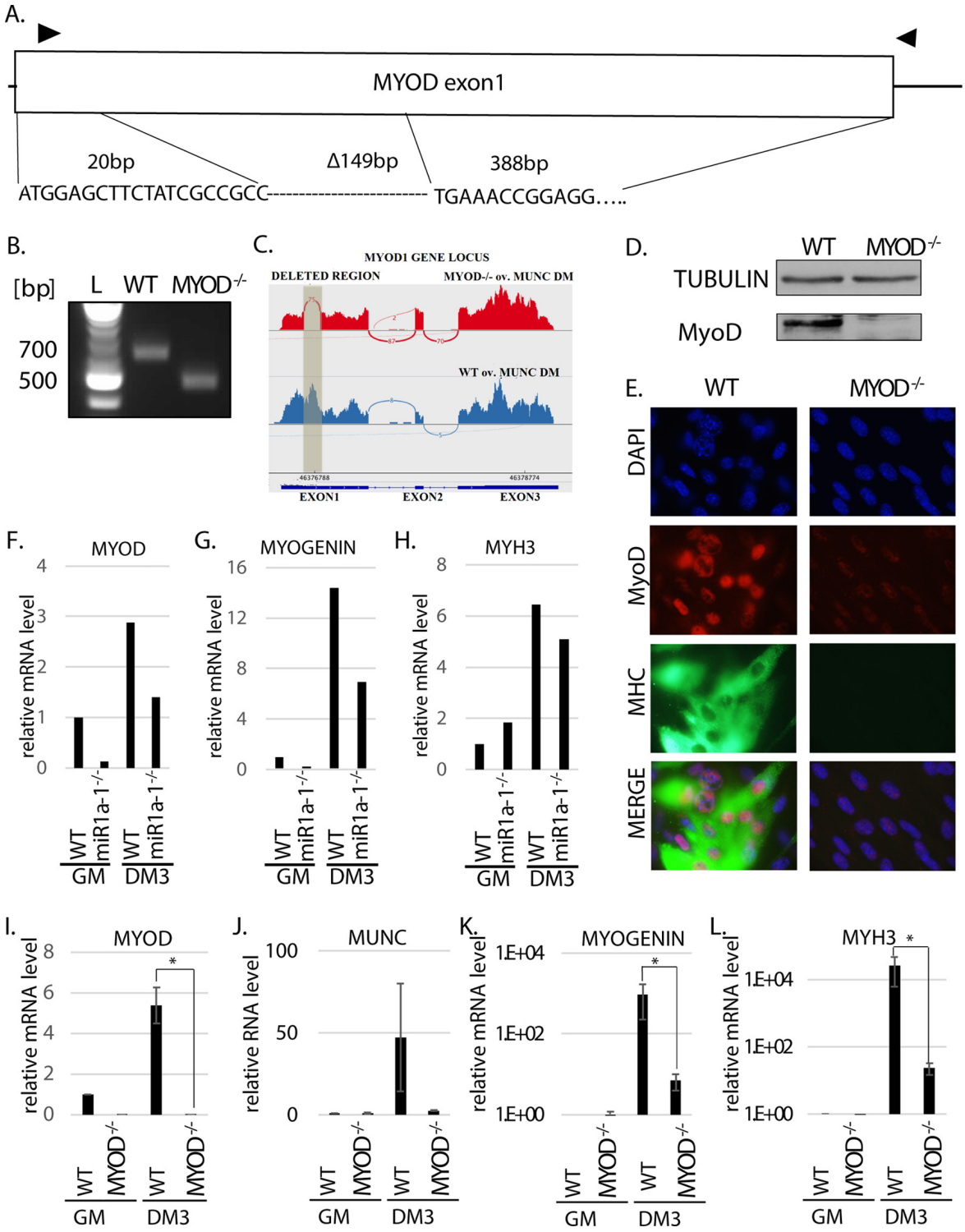


FIG 2 *MYOD* knockout decreases muscle differentiation *in vitro*. (A) Deletion of *MYOD* genomic sequence causing MyoD protein deletion. The triangles indicate the primers used for genotyping. The sequence across the deletion junction is shown and was confirmed by sequencing the genotyping PCR product from the genomic DNA of *MYOD*$^{-/-}$ cells. (B) PCR products with the genotyping primers on genomic DNA confirmed *MYOD* sequence deletion in *MYOD*$^{-/-}$ cells. The products were sequenced to confirm the deletion junction shown in panel A. The complete absence of a WT genotype band in the *MYOD*$^{-/-}$ cells confirmed that no WT allele was left. (C) RNA-Seq confirmed deletion of all alleles of MyoD in the *MYOD*$^{-/-}$ cells. A Sashimi plot of the RNA-Seq reads shows that the deleted region (shaded) in exon 1 of the *MYOD* gene is missing in *MYOD*$^{-/-}$ cell RNA. (D) Western blot analysis confirms the absence of MyoD protein in *MYOD*$^{-/-}$ cells. Tubulin served as a loading control. (E) Immunofluorescence analysis of fixed cells 3 days after differentiation (DM3). The cells were immunostained with antibodies against MyoD and MHC. DAPI (4′,6-diamidino-2-phenylindole) was used to visualize nuclei. (F to H) qRT-PCR analysis of proliferating (GM) and differentiating (DM3) C2C12 cells that were WT or miR-1a-1$^{-/-}$. Levels of *MYOD*, *MYOGENIN*, and *MYH3* mRNAs normalized to GAPDH are shown relative to that in proliferating WT cells (WT GM). (I to L) qRT-PCR analysis

(Continued on next page)

the cells, confirmed by antibodies recognizing an epitope in the C terminus of the protein (Fig. 2D).

To ensure that nonspecific effects of CRISPR-Cas9 editing or clonal selection did not impair differentiation, we also engineered a C2C12 cell with homozygous deletion of miR1a-1. This microRNA is not expected to be essential for muscle differentiation because of the presence in skeletal muscle of two other microRNAs from the same sequence family, miR-206 and miR1a-2. The miR1a-1$^{-/-}$ cells differentiated in DM3 and induced the RNAs of three myogenic factors, *MYOD*, *MYOGENIN*, and *MYH3*, almost as efficiently as WT cells (Fig. 2F to H). Thus, CRISPR-Cas9 editing or clonal selection does not impair C2C12 cell differentiation.

In contrast, the *MYOD*$^{-/-}$ cells differentiated poorly. WT cells showed the expected induction of specific myogenic transcripts after differentiation: *MYOD* (Fig. 2I), MUNC (Fig. 2J), *MYOGENIN*, and *MYH3* (Fig. 2K and L). In contrast, *MYOD*$^{-/-}$ cells with part of the *MYOD* transcript deleted (Fig. 2I) had low expression of MUNC (Fig. 2J) and nearly 100-fold less induction of *MYOGENIN* or *MYH3* RNA than WT cells (Fig. 2K and L). Note that because of the nearly undetectable levels of *MYOGENIN* or *MYH3* mRNA in C2C12 cells in GM, there is great variation in the high threshold cycle values (number of qPCR cycles after which the product becomes detectable) of these two transcripts in GM from experiment to experiment. Therefore, the fold induction in DM relative to this basal level varies greatly from experiment to experiment, even in WT cells, e.g., 14-fold versus 1,000-fold for *MYOGENIN* (Fig. 2G versus K) and 6-fold versus 30-fold for *MYH3* (Fig. 2H versus L). This is why the fold induction during differentiation shown in the figures should not be compared between experiments but should always be interpreted relative to control cells included in each experiment. Thus, we conclude that the miR1a-1$^{-/-}$ cells are almost as good as WT cells at inducing the myogenic transcripts, while the *MYOD*$^{-/-}$ cells are 100-fold weaker than WT cells at inducing the same transcripts.

Consistent with this, the *MYOD*$^{-/-}$ cells lacked MyoD and myosin heavy chain (MHC) proteins by immunofluorescence assay (the background signal is due to incomplete cutoff by the filter) in DM3 (Fig. 2E). These results agree with previous reports that *MYOD* is essential for myogenesis *in vitro* and confirm that we successfully deleted *MYOD* in the C2C12 cells.

MUNC knockout disrupts myogenesis, which is rescued by overexpression of MyoD. In parallel, we generated *MUNC*$^{-/-}$ C2C12 clones (Fig. 3A). The deletion of MUNC by CRISPR-Cas9 engineering was confirmed by PCR of genomic DNA (Fig. 3B) and Sanger sequencing of the PCR products (Fig. 3A). To confirm deletion of MUNC sequence, we performed Southern blotting of genomic DNA digested with BspHI enzyme. The digestion sites are labeled in Fig. 3A. WT cells produced a 9-kb DNA band, and *MUNC*$^{-/-}$ cells produced an 8-kb band, confirming full deletion of both alleles of MUNC (Fig. 3C). *MUNC*$^{-/-}$ cells were disabled in differentiation: *MYOD*, *MYOGENIN*, and *MYH3* RNAs were decreased at least 5-fold compared to WT DM3 cells (Fig. 3D, E, and F). To examine whether the induction of the two RNAs was rescued by addition of MyoD, we overexpressed MyoD using a doxycycline-inducible MyoD-expressing lentivirus vector (Fig. 3H). After 3 days of differentiation, this was sufficient to induce *MYOGENIN* and *MYH3* RNAs (Fig. 3G, lane 3 versus lane 2). This was accompanied by the induction of myogenin and MHC proteins (Fig. 3H) and morphological differentiation.

To ensure that the failure to differentiate seen in the knockout cells was not due to delayed kinetics of differentiation and to compare the two types of knockout cells with each other, we compared the differentiation efficiencies of WT, *MUNC*$^{-/-}$, and *MYOD*$^{-/-}$ cells over 5 days by measuring mRNA levels of myogenic factors (Fig. 4A to D). WT cells, as expected, showed a progressive increase of *MYOD* (Fig. 4A), MUNC

FIG 2 Legend (Continued)
of proliferating (GM) and differentiating (DM3) cells that were WT or *MYOD*$^{-/-}$. Levels of the indicated RNAs normalized to GAPDH are shown relative to that in proliferating WT cells (WT GM). The values represent three biological replicates and are presented as means and SEM. Statistical significance was calculated using the Wilcoxon-Mann-Whitney test. *, $P < 0.05$.

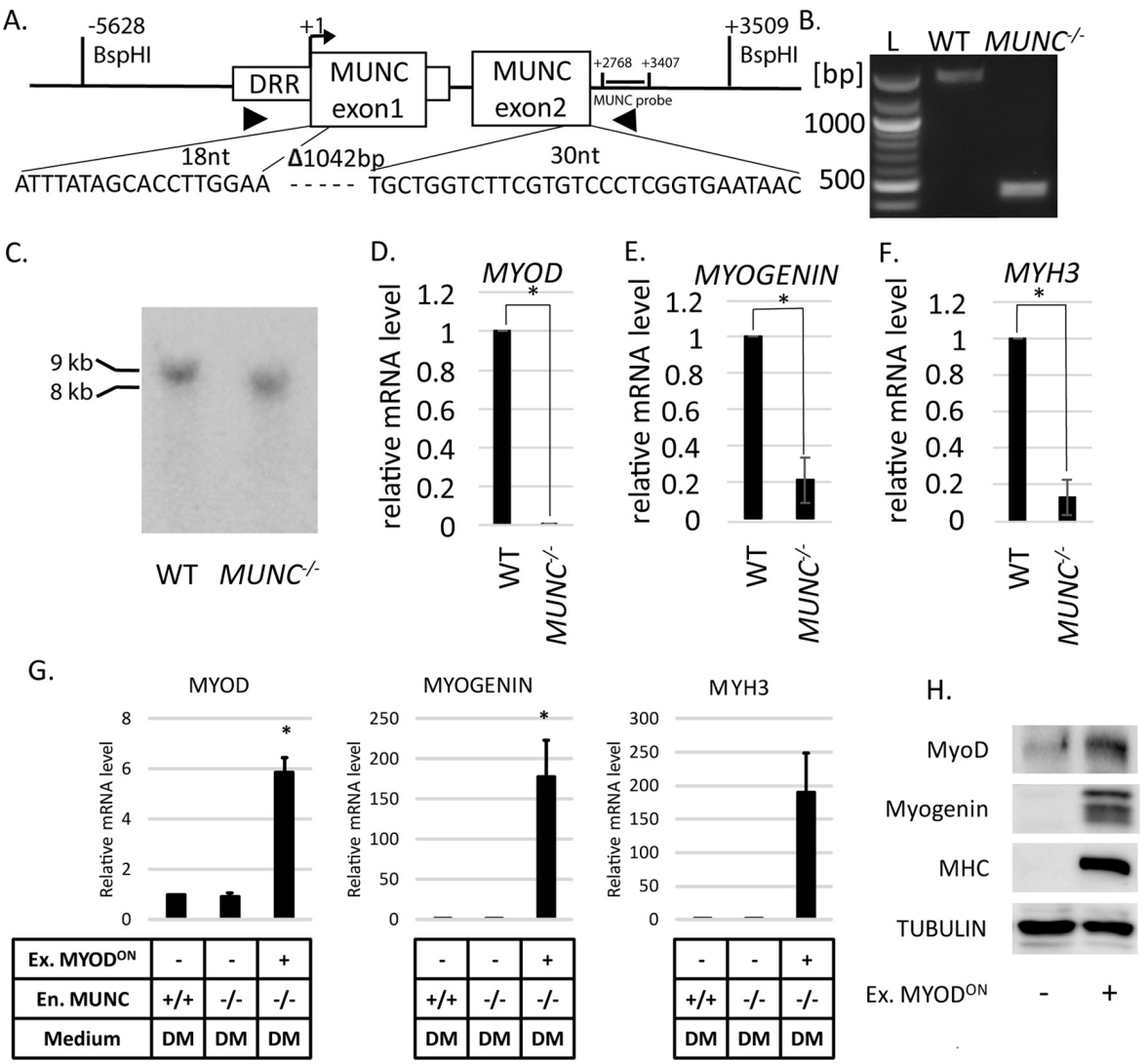


FIG 3 MUNC knockout decreases expression of *MYOGENIN* and *MYH3* RNAs, which is rescued by overexpression of MYOD. (A) The segment of MUNC genomic sequence that was deleted. The triangles indicate target sites for genotyping primers. Sequencing of the genotyping PCR products confirmed the deletion junction shown below. Locations of BspH1 restriction sites and the MUNC probe for Southern blotting are shown relative to the MUNC TSS. (B) PCR products genotyping MUNC in WT and *MUNC*$^{-/-}$ cells. (C) Confirmation of MUNC deletion by Southern blotting hybridization of BspHI-digested genomic DNA. The sizes of DNA fragments that hybridize with the MUNC probe are consistent with predicted sizes of genomic DNA from WT and *MUNC*$^{-/-}$ cells. (D to F) qRT-PCR analysis of differentiating (DM3) WT cells or *MUNC*$^{-/-}$ cells. The levels of the indicated mRNAs were normalized to GAPDH and are shown relative to WT cells. The values represent three biological replicates and are presented as means and SEM. Statistical significance was calculated using the Wilcoxon-Mann-Whitney test. *, $P < 0.05$. (G) qRT-PCR of the indicated RNAs in WT and *MUNC*$^{-/-}$ cells after 3 days in DM. The RNA levels were normalized to GAPDH and expressed relative to the level in WT cells. All the cells were transduced with lentivirus containing *MYOD*. Ex. MYOD^{ON}, lentiviral MYOD induced by doxycycline. (H) Western blotting for the indicated proteins in MUNC$^{-/-}$ cells with and without MyoD overexpression.

(Fig. 4B), *MYOGENIN* (Fig. 4C), and *MYH3* (Fig. 4D) after 1, 3, and 5 days of differentiation. *MUNC*$^{-/-}$ cells did not show any induction of myogenic mRNAs. In *MYOD*$^{-/-}$ cells, levels of RNA markers were very low compared to WT cells but were slightly induced after 5 days of differentiation. Neither mutant was able to synthesize myogenin or MHC protein (Fig. 4E), suggesting that they do not differentiate much, even though *MYOGENIN* and MUNC RNAs were induced to low levels in the *MYOD*$^{-/-}$ cells. Immunostaining cells after 5 days of differentiation showed many myotubes containing MHC in WT cells and none in either of the mutants (Fig. 4F). These results confirm that deletion of *MYOD* or MUNC equally impairs muscle differentiation.

Stable overexpression of MUNC in *MYOD*$^{-/-}$ cells induces *MYOGENIN* and *MYH3* transcripts and proteins in the complete absence of MyoD protein. We have

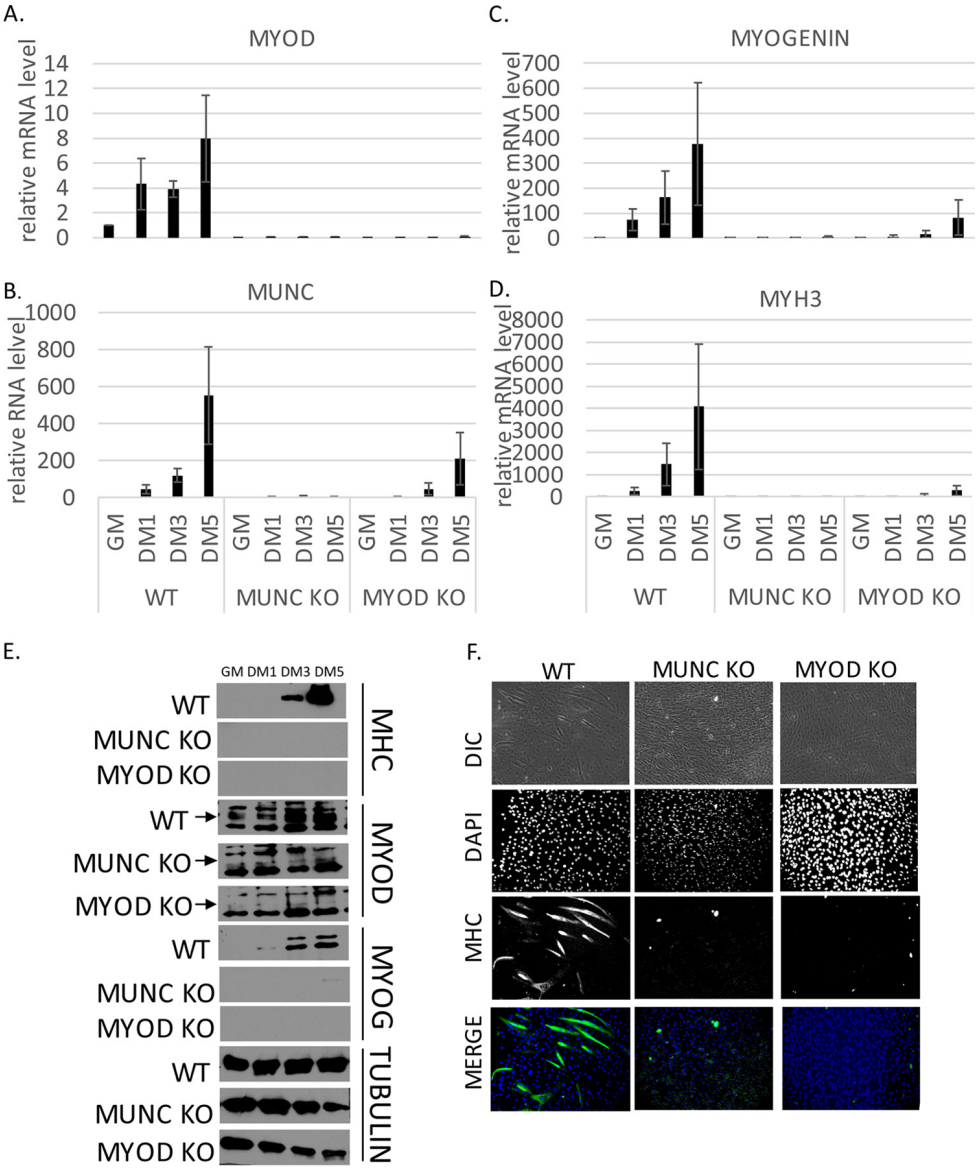


FIG 4 The time course of differentiation confirms that *MUNC*$^{-/-}$ and *MYOD*$^{-/-}$ C2C12 cells do not differentiate *in vitro*. (A to D) qRT-PCR analysis of proliferating (GM) and differentiating (DM1, DM3, and DM5) cells that were WT for MYOD and MUNC and of *MUNC*$^{-/-}$ and *MYOD*$^{-/-}$ cells. Levels of the indicated RNAs were normalized to GAPDH and are shown relative to that in proliferating WT cells (WT GM). The values represent three biological replicates and are presented as means and SEM. (E) Western blot of proliferating (GM) and differentiating (DM1, DM3, and DM5) cells that were WT for *MYOD* and for *MUNC* and of *MUNC*$^{-/-}$ or *MYOD*$^{-/-}$ cells. Protein levels for MyoD, myogenin, and MHC were measured. Tubulin served as a loading control. An arrow indicates the specific band for MyoD protein. (F) Immunofluorescence analysis of fixed cells 5 days after differentiation (DM5). Cells were immunostained with antibodies against MHC. DAPI was used to visualize nuclei. DIC, differential interference contrast.

seen the induction of *MYOGENIN* and *MYH3* RNAs in WT C2C12 cells by the overexpression of MUNC (13). In those experiments, MyoD protein was not induced any further, but it was still present, so we could not definitively say that MUNC induced these RNAs independently of MyoD. We could rule out any role of MyoD by stably overexpressing spliced MUNC in *MYOD*$^{-/-}$ C2C12 cells. Cells overexpressing MUNC (Fig. 5A) showed higher expression of *MYOGENIN* RNA in both GM (100-fold induction) and DM (10-fold induction) than control cells not overexpressing MUNC (Fig. 5B). *MYH3* RNA was also increased by 10-fold in both GM and DM (Fig. 5C). Thus, the lncRNA MUNC is able to induce *MYOGENIN* and *MYH3* RNAs in the complete absence of MyoD protein.

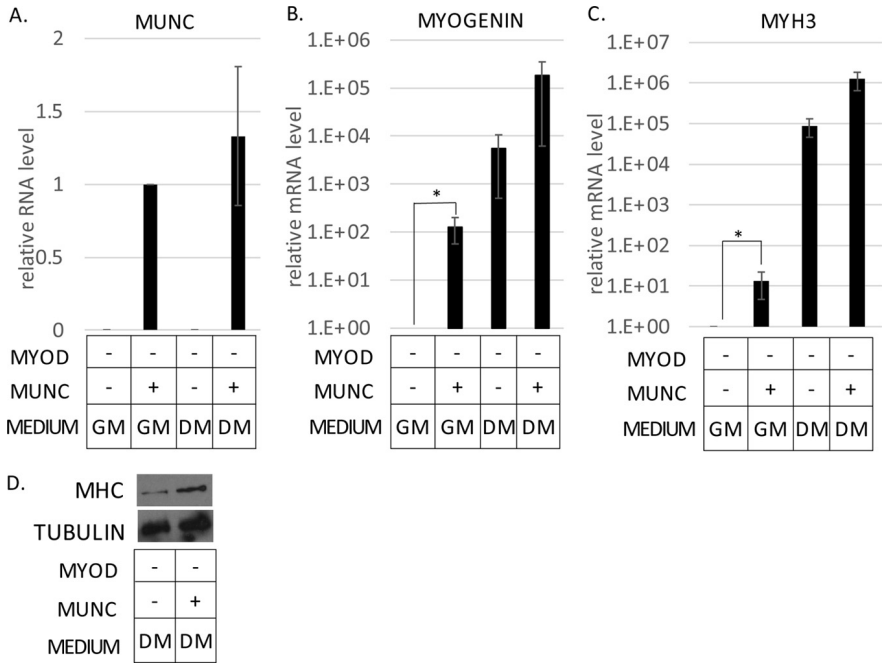


FIG 5 Stable overexpression of *MUNC* in *MYOD*$^{-/-}$ cells induces *MYOGENIN* and *MYH3* transcript levels independently of MyoD. (A to C) qRT-PCR analysis of RNAs from proliferating (GM) and differentiating (DM3) *MYOD*$^{-/-}$ cells stably transfected with vector expressing MUNC. Levels of the indicated RNAs were normalized to GAPDH and are shown relative to *MYOD*$^{-/-}$ proliferating cells (GM). +, overexpression of exogenous MUNC. The values represent three biological replicates and are presented as means and SEM. Statistical significance was calculated using the Wilcoxon-Mann-Whitney test. *, $P < 0.05$. (D) Western blot analysis of MHC protein levels in *MYOD*$^{-/-}$ cells overexpressing MUNC under DM3 conditions. Tubulin was used as a loading control.

In WT C2C12 cells, the induction of *MYOGENIN* or *MYH3* RNA by MUNC was not accompanied by the induction of the two proteins (13). However, the situation was slightly different in the *MYOD*$^{-/-}$ cells (Fig. 5D). Even though myogenin protein was not induced under GM or DM conditions, MHC protein was slightly induced upon MUNC overexpression only under DM conditions. The *MYOD*$^{-/-}$ cells overexpressing MUNC did not show any morphological signs of differentiation. The induction of MHC protein by MUNC in *MYOD*$^{-/-}$ cells can be explained by the *MYH3* RNA reaching a threshold value in DM for the resulting protein to be detectable. However, the lack of MyoD protein still prevents the induction of myogenin protein or morphological differentiation in DM.

MyoD and MUNC cooperate to induce *MYOGENIN* and *MYH3* RNAs but fail to promote differentiation of *MYOD*$^{-/-}$ cells in DM. The RNA-Seq results we describe below suggest some cooperation between MUNC and MyoD in inducing *MYOGENIN* and *MYH3* RNAs. MUNC overexpression is better in WT cells than in *MYOD*$^{-/-}$ cells at inducing *MYOGENIN* (5-fold) and *MYH3* (12-fold). We therefore wanted to test whether MyoD synergizes with MUNC for the induction of these two genes. First we tested the maximum extent to which MyoD protein restoration in *MYOD*$^{-/-}$ cells would induce *MYOGENIN* and *MYH3* RNAs by lentivirus-mediated doxycycline-inducible overexpression of MyoD (Fig. 6A and D). Exogenous MyoD induced *MYOGENIN* RNA and protein and *MYH3* RNA (Fig. 6B, C, and D).

In order to see cooperation between MyoD and MUNC for the induction of *MYOGENIN* and *MYH3* RNAs, *MYOD*$^{-/-}$ cells stably overexpressing MUNC were transiently transfected with a plasmid vector expressing *MYOD* (Fig. 6E and I). Relative to control cells, *MYOGENIN* RNA was induced 3-fold by MyoD alone, 4-fold by MUNC alone, and 6-fold by both MyoD and MUNC (Fig. 6G). *MYH3* RNA was similarly induced 3-fold by MyoD alone, 4.5-fold by MUNC alone, and 6-fold by both MyoD and MUNC (Fig. 6H). Although MUNC plus MyoD genes induced more RNA than either gene alone, the differences did

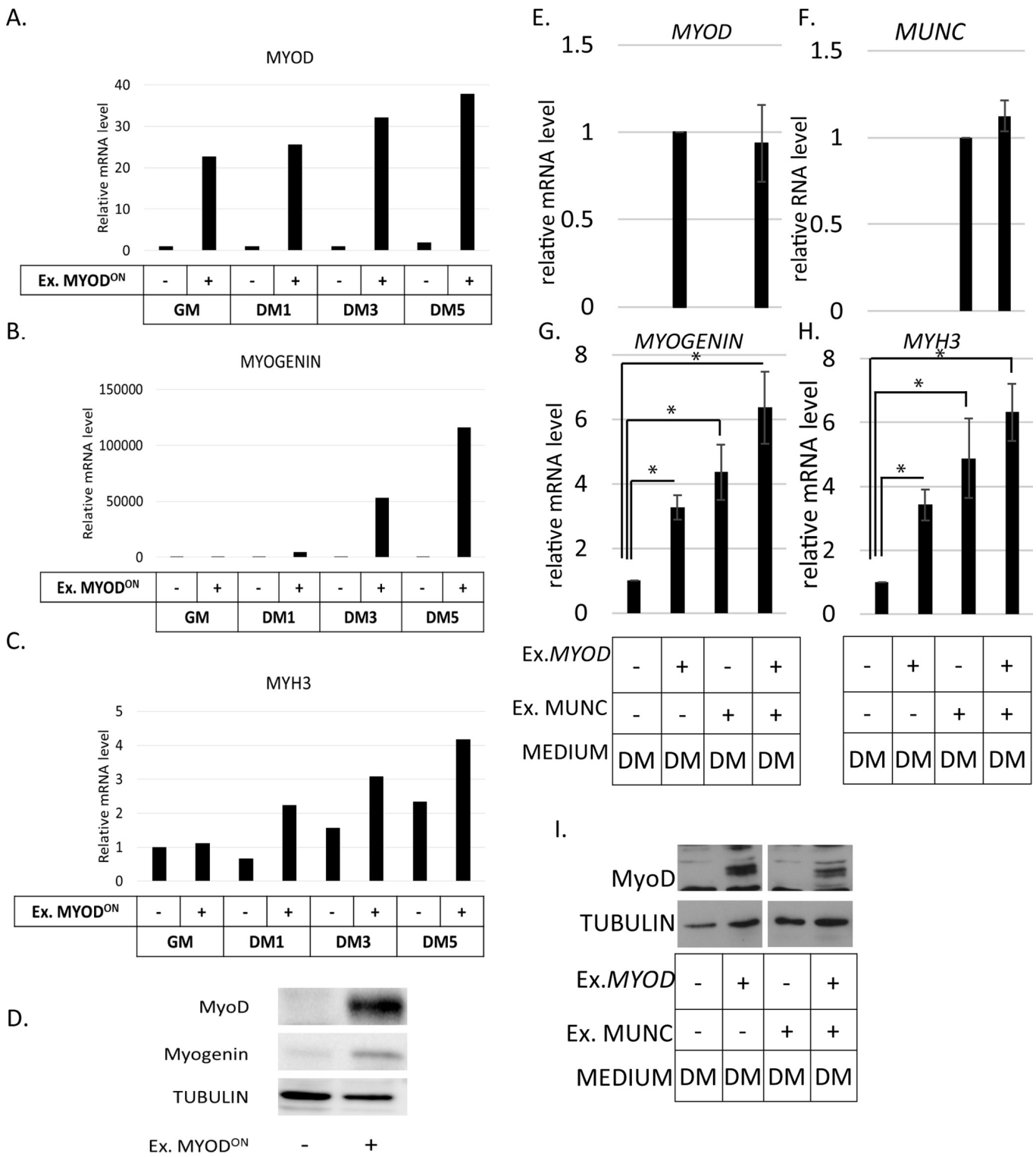


FIG 6 Expression of MYOD partially rescued the *MYOD*$^{-/-}$ cell phenotype but did not significantly stimulate the induction of *MYOGENIN* or *MYH3* by MUNC. (A to C) qRT-PCR analysis of *MYOD*$^{-/-}$ cells with or without doxycycline-mediated overexpression of exogenous *MYOD* in proliferating (GM) and differentiating (DM1, DM3, and DM5) cells as indicated on the *x* axes. Levels of expression were measured for *MYOD* (A), *MYOGENIN* (B), and *MYH3* (C) mRNAs and normalized to GAPDH and are shown relative to *MYOD*$^{-/-}$ cells without overexpressed MYOD in GM. (D) Western blot showing exogenous MyoD and myogenin proteins induced in *MYOD*$^{-/-}$ cells when exogenous MyoD protein is induced by doxycycline. Tubulin was used as a loading control. (E to H) qRT-PCR analysis of *MYOD*$^{-/-}$ cells stably overexpressing MUNC, transiently overexpressing exogenous *MYOD*, and differentiated for 2 days. Levels of expression were measured for *MYOD* (E), MUNC (F), *MYOGENIN* (G), and *MYH3* (H) mRNAs. The data were normalized to the GAPDH expression level and are shown relative to control cells. The values represent three biological replicates and are presented as means and SEM. Statistical significance was calculated using a Wilcoxon-Mann-Whitney test. *, $P < 0.05$. (I) Western blot analysis showing induction of exogenous MyoD protein in *MYOD*$^{-/-}$ cells when transiently transfected with MYOD. Tubulin was used as a loading control.

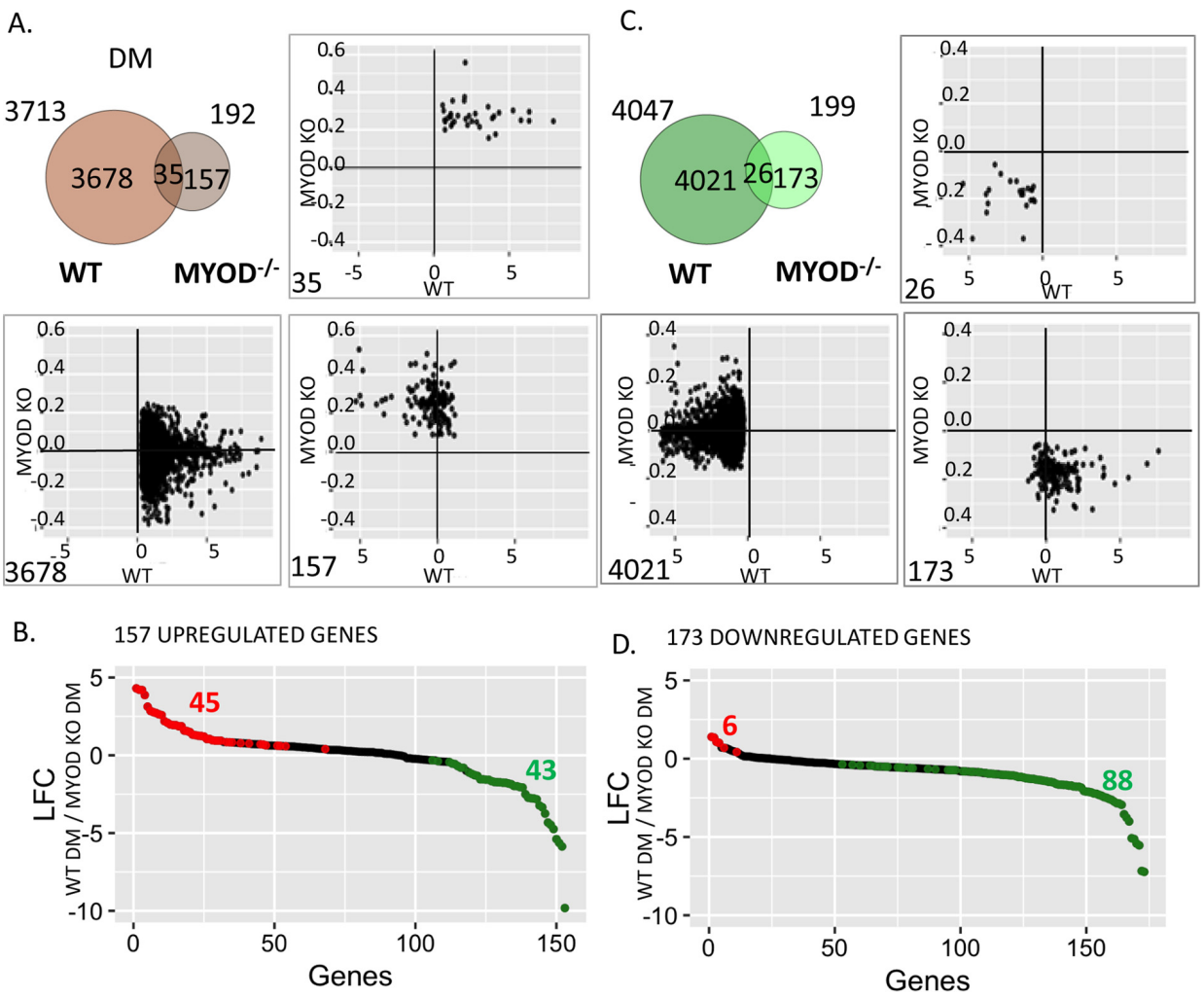


FIG 7 MUNC overexpression regulates many cellular genes in the complete absence of MyoD protein. (A) Venn diagram representing overlap of genes that are upregulated upon MUNC overexpression in WT or *MYOD*$^{-/-}$ cells at DM3. The scatter plots show how MUNC overexpression regulates ($\log_2$ fold change in cells overexpressing MUNC relative to control cells transfected with the empty vector) the three classes of genes in WT cells and *MYOD*$^{-/-}$ cells. (B) The 157 genes upregulated by MUNC only in *MYOD*$^{-/-}$ cells were examined to see if they were induced or repressed by MyoD. The plots represent $\log_2$ fold changes of genes on DM3 in WT versus *MYOD*$^{-/-}$ cells. The red and green dots represent genes that were induced or repressed in WT cells (induced or repressed by the presence of MyoD protein); $P < 0.05$. (C) Same as panel A, except for genes downregulated upon MUNC overexpression in WT or *MYOD*$^{-/-}$ cells. (D) Same as panel B, except for 173 genes from panel C that were downregulated by MUNC only in *MYOD*$^{-/-}$ cells.

not reach statistical significance in DM. Interestingly, in GM, where the basal levels of *MYOGENIN* and *MYH3* RNAs are lower, we saw statistically significant additive stimulation of the two RNAs upon coexpression of MUNC and *MYOD* (not shown), but the levels of *MYOGENIN* and *MYH3* RNAs did not reach the levels seen during normal differentiation, and there was no morphological differentiation. Thus, transient expression of MyoD, expressed from heterologous sites, had a weak additive effect with MUNC to induce more *MYOGENIN* or *MYH3* RNAs, but it was not statistically significant and was insufficient to promote any morphological differentiation in *MYOD*$^{-/-}$ cells.

MUNC overexpression regulates genes both in cooperation with MyoD and in the complete absence of MyoD. MUNC induced *MYOGENIN* and *MYH3* even in *MYOD*$^{-/-}$ cells, suggesting that it can act independently of MyoD. However, MUNC also induced *MYOD*, suggesting that the two genes could cooperate with each other in regulating gene expression. To determine how many genes are regulated by MUNC independently of MyoD and how many in cooperation with MyoD, we examined the global RNA changes produced by MUNC overexpression in WT cells and *MYOD*$^{-/-}$ cells after 3 days of differentiation (DM3) (Fig. 7A and C). The Venn diagram in Fig. 7A shows

that 3,678 genes were induced by MUNC only in WT cells but not in *MYOD*$^{-/-}$ cells, suggesting that there is a large fraction of genes that are induced by MUNC only in the presence of MyoD. This could be either because MyoD stimulates these genes and MUNC increases MyoD protein expression or because there is cooperation between MUNC and MyoD (or a MyoD-induced factor) at these promoters. There were 35 genes similar to *MYOGENIN* and *MYH3* that were induced by MUNC in the presence or absence of MyoD and 157 genes that were induced by MUNC in *MYOD*$^{-/-}$ cells but not in WT cells. These two groups clearly show that MUNC can regulate the expression of 192 genes independently of MyoD protein.

The scatter plots in Fig. 7A show how individual genes in each of these three groups behave upon MUNC overexpression in WT cells and in *MYOD*$^{-/-}$ cells. The 35 genes that were induced in both types of cells (upper-right plot), were less induced in the absence of MyoD. The 3,678 genes that were induced by MUNC exclusively in WT cells (lower-left plot) were mostly unaffected in the *MYOD*$^{-/-}$ cells ($\log_2$ fold change from 0.2 to -0.2), though there were a few that were repressed by MUNC in the absence of MyoD. Surprisingly, of the 157 genes that were induced by MUNC exclusively in the *MYOD*$^{-/-}$ cells (lower-right plot), a large number were repressed by MUNC in WT cells, suggesting that the presence of MyoD reverses the direction of change produced by MUNC.

The last observation raises the possibility that MyoD induced by MUNC overexpression is responsible for the repression of the 157 genes. If that is the case, all 157 genes would be expressed less in WT cells than in *MYOD*$^{-/-}$ cells even without overexpressing MUNC (Fig. 7B). Out of these 157 genes, expression of only 43 was lower in WT cells than in *MYOD*$^{-/-}$ cells ($P < 0.05$), suggesting that they were repressed by MyoD and so could be repressed by MUNC through the induction of MyoD. However, in the absence of MyoD, MUNC induces these genes, providing further support for the hypothesis that MUNC regulates many genes completely independently of the MyoD protein. The 45 genes at the left end of the plot in Fig. 7B were induced by the presence of MyoD, so their repression by MUNC in WT cells cannot be explained by postulating an indirect effect through the induction of MyoD by MUNC.

Turning to genes repressed by MUNC under differentiating conditions (Fig. 7C), we found 4,021 genes that were repressed by MUNC only in the presence of MyoD. MUNC either represses these genes indirectly through the induction of MyoD or cooperates with MyoD (or some MyoD-induced factor) at their promoters. We analyzed chromatin immunoprecipitation sequencing (ChIP-seq) data available for MyoD protein in C2C12 cells to determine whether the repressed genes had MyoD binding sites near their transcription start sites (12). Forty-seven percent of the repressed genes were closest to (nearest neighbors to) a MyoD binding site. Thus, at least 53% of the genes repressed by MUNC are repressed indirectly by cooperation with some factor present when MyoD is present, but not MyoD itself. Looked at another way, only 47% of the 4,021 genes are repressed by MyoD alone in MyoD-converted mouse embryonic fibroblasts (MEFs) and so could be repressed by MUNC through the induction of MyoD (26). However, all these genes do not contain MyoD ChIP sites and so may be repressed indirectly by factors induced by MyoD. The data suggest that many genes are repressed by overexpressed MUNC in MyoD^{+} cells in direct or indirect cooperation with MyoD.

Twenty-six genes were repressed by MUNC in the presence or absence of MyoD, and 173 genes were repressed by MUNC only in the absence of MyoD, again showing evidence of MUNC activity independent of MyoD protein. The scatter plot in Fig. 7C, lower right, suggests that the 173 genes repressed by MUNC in *MYOD*$^{-/-}$ cells include many genes that are paradoxically upregulated by MUNC in WT cells. Among these, the plot in Fig. 7D identifies 6 genes that are induced by the presence of MyoD and so might be induced by MUNC in WT cells through the induction of MyoD. However, in the absence of MyoD, MUNC independently acts on the same genes and represses them. Figure 7D also identifies 88 genes that are repressed by MyoD (in a comparison of WT and *MYOD*$^{-/-}$ cells). These genes are repressed by MUNC in the absence of MyoD (they are among the 173 genes in Fig. 7C), and yet overexpression of MUNC in WT *MYOD*$^{+}$

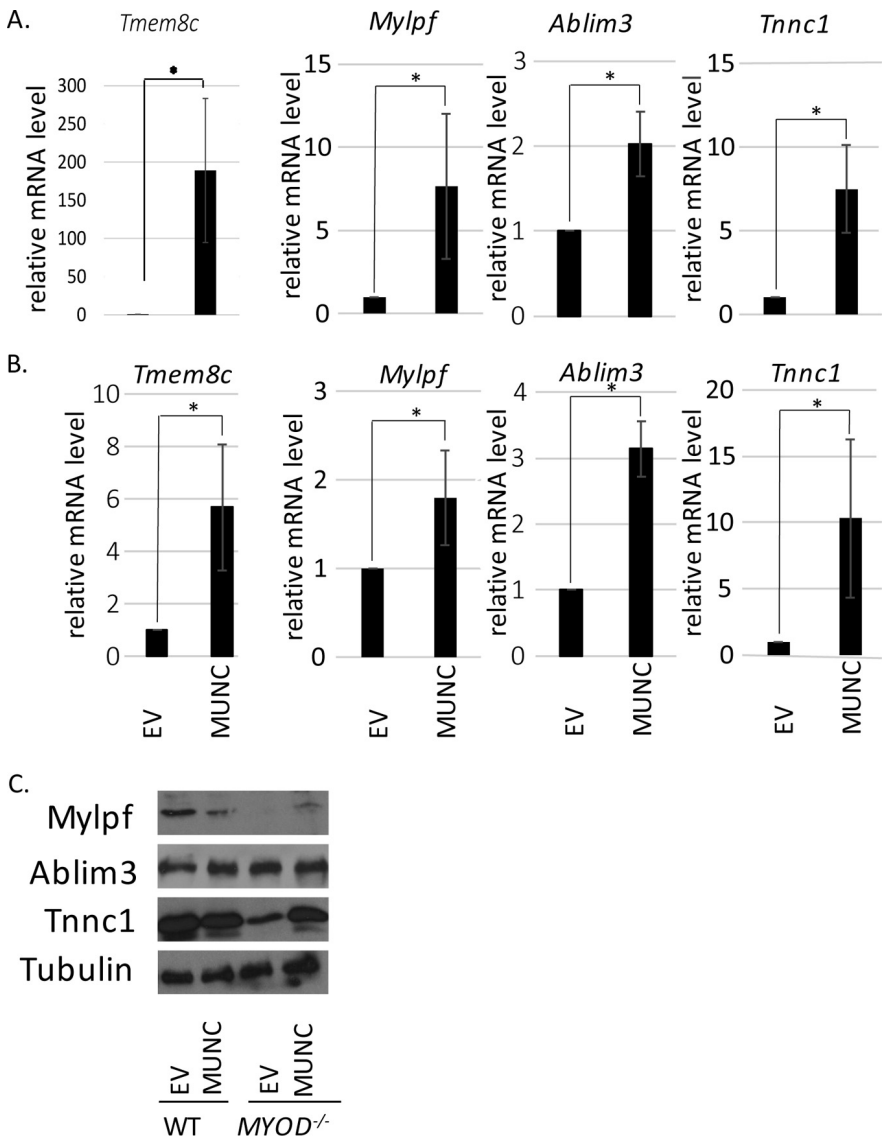


FIG 8 Other myogenic genes and proteins are induced by MUNC in *MYOD*$^{-/-}$ cells. (A and B) qRT-PCR confirmation of genes upregulated upon MUNC overexpression. Shown is analysis of WT cells (A) and *MYOD*$^{-/-}$ cells (B) under differentiating conditions. The data were normalized to the GAPDH expression level and are shown relative to control cells (EV). The values represent three biological replicates and are presented as means and SEM. Statistical significance was calculated using the Wilcoxon-Mann-Whitney test. *, $P < 0.05$. (C) Western blots of protein products of genes analyzed in panels A and B.

cells did not lead to their repression (Fig. 7C, lower right scatter plot), suggesting that MyoD and MUNC do not act additively on these promoters.

Collectively, these results suggest that MUNC and MyoD cooperate to regulate thousands of genes but that there are a few hundred genes that are regulated by MUNC in the complete absence of MyoD protein, consistent with our hypothesis that MUNC is not merely an eRNA whose only role is to induce *MYOD* transcription. Additionally, we observed a group of genes that were regulated by MyoD and MUNC in opposite directions, which also suggests independence of action. Finally, this is the first evidence that overexpressed MUNC can also repress thousands of cellular genes.

Confirmation of induction of genes by MUNC in *MYOD*$^{-/-}$ cells. We selected 4 of the 35 genes besides *MYOGENIN* and *MYH3* that are induced by MUNC in WT and *MYOD*$^{-/-}$ C2C12 cells (Fig. 7A) to confirm the induction by quantitative reverse transcription (qRT)-PCR in WT cells (Fig. 8A) and in *MYOD*$^{-/-}$ cells (Fig. 8B). We focused on genes whose products are functionally and structurally connected to skeletal muscle

function: *Tmem8c*, a gene coding for Myomaker, a protein essential for fusion of embryonic and adult myoblasts; *Mylpf*, a gene coding for the regulatory light chain of striated muscle (27); *Ablim3*, encoding a protein that binds strongly to F-actin, suggesting its role as a scaffold for actin cytoskeleton signaling (28); and *Tnnc1*, a gene coding for troponin C, a part of the troponin complex, a structural complex responsible for muscle contraction (29). All four genes were induced by MUNC in WT and *MYOD*$^{-/-}$ C2C12 cells.

We next checked the levels of protein products from *Mylpf*, *Ablim3*, and *Tnnc1* (Fig. 8C). *Tmem8c* was not studied because there are no suitable commercial antibodies (Abs) available for immunoblotting. The *MYOD*$^{-/-}$ cells in DM expressed low levels of myosin light chain (*Mylpf* product) and troponin C1 (*Tnnc1* product), allowing us to detect induction of these proteins when MUNC was overexpressed and induced the corresponding RNAs. We suggest that the levels of these proteins are regulated posttranscriptionally so that further protein induction is not seen when the protein levels are already high (as in WT cells with empty vector), even though the RNAs are induced by MUNC in WT cells.

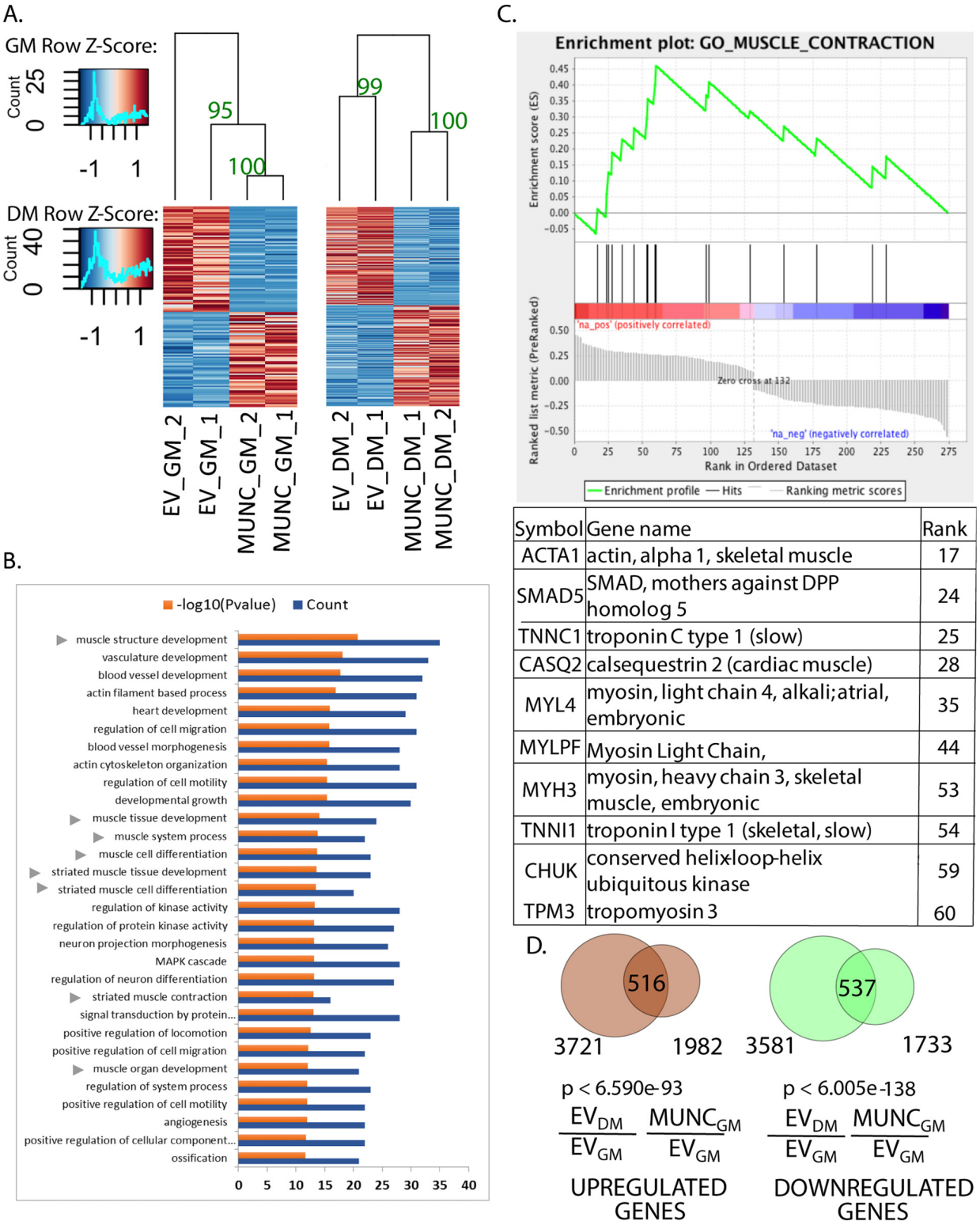
MUNC regulates muscle-related genes in *MYOD*$^{-/-}$ cells. We first tested the reproducibility of the gene expression changes seen with MUNC overexpression independently of MyoD. Hierarchical clustering of the differentially expressed genes in *MYOD*$^{-/-}$ C2C12 cells in both GM and DM3 showed that the pattern of gene expression changes was preserved in two independent experiments (Fig. 9A). Gene ontology (GO) terms that were enriched among the genes regulated by MUNC in DM3 in the *MYOD*$^{-/-}$ cells indicated that many of them are associated with skeletal muscle development and muscle structure (Fig. 9B). Fewer genes involved in skeletal muscle development and structure are regulated by MUNC in GM in *MYOD*$^{-/-}$ cells (GO term enrichment analysis for GM not shown). Therefore, DM likely induces factors independent of MyoD that cooperate with MUNC to regulate many myogenic genes.

To determine the most significant molecular pathway regulated by MUNC in the absence of MyoD, we performed a gene set enrichment analysis (GSEA) on the genes differentially regulated upon MUNC overexpression in *MYOD*$^{-/-}$ cells in DM3. The plot in Fig. 9C shows significant enrichment of genes involved in muscle contraction among the genes induced by MUNC. The table below the plot lists the top 10 genes contributing to the enrichment score for muscle contraction GO terms, which are mainly muscle structure protein-coding genes.

As discussed above, the MyoD-independent activity of MUNC is more myogenic in DM than in GM, but we wanted to test whether the global change in gene expression induced by MUNC in WT C2C12 cells in GM is similar to that seen when the same cells undergo differentiation in DM. A total of 1,982 genes were induced and 1,733 genes were repressed by MUNC in WT cells growing in GM. When these genes were compared with the genes that were induced or repressed upon differentiation of WT C2C12 cells, a highly significant number of genes were found to overlap (Fig. 9D). This result suggests that MUNC overexpression alone in GM is able to push C2C12 cells in the direction of myogenic differentiation, although of course, MUNC overexpression alone is not as potent as the differentiation induced by moving cells from GM to DM.

DISCUSSION

The first question this paper answers is whether MUNC is an lncRNA that has functions independent of acting as an eRNA for *MYOD* (Fig. 10). Recent reports suggest that long noncoding RNAs derived from enhancer loci directly regulate the expression level of neighboring genes by a *cis*-acting mechanism (2). p53-bound enhancer regions produce eRNAs that regulate the transcription of adjacent genes, as shown by reporter assays and RNA Pol II ChIP assay (30). Additional examples are activating noncoding RNAs (ncRNAs), ncRNA-a3 and ncRNA-a7, whose depletion decreases RNA Pol II abundance at adjacent genes, as well as the recruitment of Mediator to the adjoining promoter (46). Estrogen receptor alpha (ERα)-inducible enhancer RNAs are functionally important for the expression of their target genes and are crucial for proper chromatin



Symbol	Gene name	Rank
ACTA1	actin, alpha 1, skeletal muscle	17
SMAD5	SMAD, mothers against DPP homolog 5	24
TNNC1	troponin C type 1 (slow)	25
CASQ2	calsequestrin 2 (cardiac muscle)	28
MYL4	myosin, light chain 4, alkali;atrial, embryonic	35
MYLPF	Myosin Light Chain,	44
MYH3	myosin, heavy chain 3, skeletal muscle, embryonic	53
TNNI1	troponin I type 1 (skeletal, slow)	54
CHUK	conserved helix-loop-helix ubiquitous kinase	59
TPM3	tropomyosin 3	60

FIG 9 MUNC globally regulates many muscle-related genes in *MYOD*$^{-/-}$ cells. (A) Heat maps showing clustering of samples based on differentially regulated genes upon MUNC overexpression under proliferating conditions (GM) (left) and under differentiating conditions (DM) (right) in *MYOD*$^{-/-}$ cells. There were two biological replicates for each condition. Bootstrap values based on 1,000 repetitions are shown near the corresponding branches. (B) Top 30 significant gene ontology terms enriched in differentially expressed genes in DM upon MUNC overexpression in *MYOD*$^{-/-}$ cells. The arrowheads indicate gene terms related to skeletal muscle development and regeneration. (C) Enrichment plot from GSEA showing that the gene set involved in muscle contraction is enriched among differentially regulated genes upon MUNC overexpression in *MYOD*$^{-/-}$ cells in DM ($P < 0.01$). The table lists the top 10 genes contributing to enrichment scores for muscle contraction GO terms. (D) Venn diagrams representing overlap between differentially expressed genes upon differentiation of control cells (EV DM/EV GM) versus differentially expressed genes upon MUNC overexpression under proliferating conditions (MUNC GM/EV GM) in WT cells.

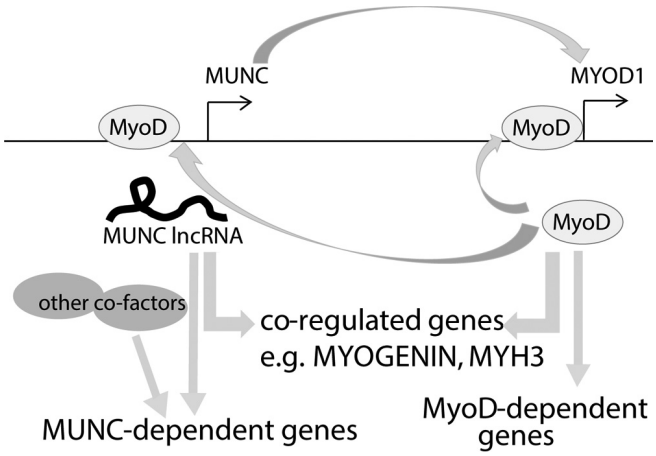


FIG 10 Schematic showing that MUNC and *MYOD* positively regulate each other and coregulate many genes but also regulate many genes independently of each other.

looping between enhancer loci and target gene bodies, which facilitates interactions between chromatin modifiers and transcription machinery (31). It was suggested that MUNC, coded by DRR genomic sequence, acts primarily as an enhancer RNA (12), inducing transcription of *MYOD*, but also induced *MYOGENIN* in *trans* (perhaps through the induction of MYOD). We now present data showing that MUNC positively regulates different myogenic genes, not only *MYOD*, and that it has many target genes that are regulated by MUNC overexpression in the complete absence of MyoD protein. The fact that specific sequence and structural elements of MUNC are necessary for the induction of *MYOD*, *MYOGENIN*, or *MYH3* argues that the mere act of transcription or splicing of MUNC is not sufficient for its activity, as has been suggested for eRNAs (2). In addition, the structure-function studies show that even in WT cells, different parts of MUNC stimulate *MYOD*, *MYOGENIN*, and *MYH3* RNAs to different extents that are not correlated with each other, something that would have been expected if all of MUNC's actions were through the induction of *MYOD* RNA and protein. These results suggest that MUNC is both a classical eRNA that induces transcription of the adjoining *MYOD* RNA and also a *trans*-acting lncRNA that has actions independent of *MYOD* induction.

This result raises the possibility that there are other eRNAs that also act as lncRNAs. So far, reports suggest that eRNAs are not spliced, that transcription from the enhancer region is bidirectional, and that transcriptionally active enhancers are tagged with H3K4me1 rather than H3K4me3 marks. Enhancer RNAs are also usually much shorter than lncRNAs (32). We know from this report and our previous study (13) that MUNC is spliced, that the predominant stable transcript at the DRR locus is in the direction of MUNC, and that the DRR genomic locus during muscle differentiation acquires H3K4me3 marks. We hypothesize that eRNAs with similar features may have dual actions as an eRNA (enhancing the transcription of the adjoining gene) and as an lncRNA, which executes functions independent of its nearby neighbor.

The next question is whether MUNC lncRNA acts through the expression of an encoded micropeptide. There are growing reports that some lncRNAs code for functional micropeptides of even 30 amino acids. The most recent examples are micropeptides described by Olson and colleagues, which by interaction with SERCA regulate calcium signaling in muscle (33, 34) and nonmuscle (35) cells. Additionally, it was shown that one genomic locus may produce both a functional micropeptide, MLN, and a functional lncRNA, linc-RAM, working independently of each other (22). Spliced MUNC transcript could code for three such micropeptides unrelated in sequence to each other (underlined in red in Fig. 1A). However, the structure-function studies on MUNC rule out the possibility that MUNC's lncRNA-like function is due to any of the three putative micro-open reading frames (ORFs) in MUNC and suggest instead that the sequence and the folding of the RNA fragments are important for their function.

Both MUNC and MyoD are promyogenic factors, raising the question of whether they are additive with each other and whether they ever act in opposite directions. Our results suggest that, indeed, MUNC and MyoD cooperate to regulate many genes. However, there is a clear subset of genes that are regulated by MUNC in the complete absence of MyoD protein. Additionally, we observed a group of genes that are regulated by MyoD and MUNC in opposite directions, which suggests that the two factors may work in some pathways as antagonists.

The lack of MUNC or MyoD disables differentiation *in vitro*. The weak induction of some myogenic transcripts, like *MYOGENIN* and *MYH3*, when *MYOD*$^{-/-}$ cells are moved to DM (Fig. 2 and 4) is very slight relative to what is seen with WT cells. Overexpression of MUNC in *MYOD*$^{-/-}$ cells in DM induces the *MYOGENIN* RNA 4- to 10-fold (Fig. 5B and 6G), while overexpression of MyoD in the same cells induces *MYOGENIN* RNA 50,000-fold (Fig. 6B, DM3), suggesting that overexpressed MUNC cannot completely compensate for the absence of MyoD. Conversely, overexpression of MyoD in *MUNC*$^{-/-}$ cells stimulated *MYOGENIN* and *MYH3* RNAs and proteins quite effectively (Fig. 3H to I). These results suggest that at loci like *MYOGENIN*, MUNC can partly compensate for lack of MyoD and vice versa, consistent with independent modes of action of MUNC and MyoD.

When MUNC was expressed stably and MyoD was expressed transiently together in the *MYOD*$^{-/-}$ cells, there was weak additive induction of *MYOGENIN* or *MYH3* RNA in DM (and more so in GM). This was insufficient to allow differentiation of the cells. Even when MyoD protein was expressed at a high level in *MYOD*$^{-/-}$ cells (Fig. 6A to D), we saw induction of *MYOGENIN* RNA and protein, but not enough to permit differentiation.

In our previous report (13), overexpression of MUNC induced the expression of three genes, *MYOD*, *MYOGENIN*, and *MYH3*, so we focused on MUNC as a positive factor for gene expression. The genomewide analysis of genes regulated by MUNC in WT cells and in *MYOD*$^{-/-}$ cells presents a more complicated picture where in both types of cells MUNC induces and represses a large number of genes. MyoD, similarly, was initially thought to be a transcriptional factor that positively regulated expression of its target genes. However, it has since been recognized that MyoD also plays a role as a repressor of transcription, in cooperation with histone deacetylase 1 (HDAC1). For example, in proliferating myoblasts, MyoD binds to the promoter region of *MYOGENIN* to recruit HDAC1 and to suppress transcription (36). After serum withdrawal, MyoD changes its interaction partners to P/CAF and activates transcription of *MYOGENIN* (36). Another study showed that MyoD can repress c-Jun-mediated activation of genes linked to an AP-1 site in C2 cells (37). Thus, MyoD may repress specific gene promoters and MUNC may cooperate with such repression. It has also been proposed that MyoD can interact with chromatin-looping proteins, such as CTCF, to disrupt repressive loops, thus inducing transcription from specific genomic regions (38). Thus, there are different, independent mechanisms by which MyoD regulates its targets. Similarly, we propose that MUNC interacts with different cellular factors to induce or repress different targets and that the induction and repression functions are sometimes MyoD dependent and sometimes not.

Although one important conclusion of this paper is that MUNC can act independently of MyoD and sometimes in the opposite direction to MyoD, it is clear that there are many functional interactions between the two promyogenic factors. For example, MyoD promotes the transcription of MUNC (as evidenced by the decrease of MUNC in the *MYOD*$^{-/-}$ cells), and MUNC promotes the expression of MyoD. In addition, there are many genes that are regulated in the same direction by MUNC (in the presence or absence of MyoD) and by MyoD. Our future goal is to describe how MUNC and MyoD cooperate on the genes that they both induce or repress. Although we have failed to detect any direct physical interaction between MyoD and MUNC, we cannot yet rule out this possibility. Transient and weak interactions between MyoD and MUNC may be functionally important but difficult to show. In addition, MyoD interacts with numerous proteins to build whole complexes that regulate the expression of target genes, and MUNC may interact with and activate another protein from such a complex or may

function as a scaffold, helping to maintain stability of interaction between transcriptional factors and chromatin remodelers.

A related goal is to describe how MUNC acts on many genes independently of MyoD (Fig. 10). We have to identify cellular proteins that interact with MUNC independently of MyoD. The MUNC-overexpressing *MYOD*$^{-/-}$ cells will be very important for such a search. As a nuclear transcript, MUNC may interact with chromatin modifiers, transcription factors, or repressors on the chromatin. Thus, we plan to examine whether we can identify specific genomic sites at which MUNC associates with the chromatin or alters the chromatin landscape without stable association with the chromatin.

An important possibility is that the MyoD-related proteins Myf5, myogenin, and MRF4 act as cofactors for MUNC. MyoD and Myf5 play redundant roles in skeletal muscle differentiation: mice with deletion of either gene remain alive and healthy, but double-knockout pups die shortly after birth (39). Studies on double-knockout mice showed that each of the factors is essential for proper development of different parts of the musculature (40). Expression of these transcription factors during development is temporally regulated: *MYF5* transcript is evident at 7.5 days postcoitum (dpc) *MYOGENIN* at 8.5 dpc, and *MYOD* at 9.5 dpc (41). In our previous study, MUNC was induced between days 11 and 15 of embryonic development, which suggested a role at later points in development, when Myf5, myogenin, and MyoD are all present. We also showed that MUNC is induced during differentiation of myoblasts, with its abundance being very low in undifferentiated C2C12 cells in proliferating medium (13). One argument against Myf5 being a cofactor for MUNC is that Myf5 does not induce myogenic gene transcription as robustly as MyoD (26) or MUNC (Fig. 7A). MUNC could also work with myogenin, another transcription factor, which is strongly induced during differentiation and whose expression is itself MUNC dependent. Identifying the cofactors that assist MUNC activity will be another important area of future research.

MATERIALS AND METHODS

Cell culture. The C2C12 mouse myoblast cell line was supplied by the American Type Culture Collection (ATCC). C2C12 cells were cultured in Dulbecco's modified Eagle's medium (DMEM)–high-glucose medium (GE Healthcare Life Sciences Co.) with 10% fetal bovine serum (FBS) (Life Technologies Co.); when differentiating, serum was switched to 2% horse serum (GE Healthcare Life Sciences Co.)

Knockout strategy. CRISPR protocol with minor changes was followed to achieve deletion of a part of the *MYOD* gene (25). Briefly, single guide RNAs (sgRNAs) were designed using the CRISPR DESIGN tool (http://crispr.mit.edu/). Cells were cotransfected with vectors coding for Cas9 (the vectors were obtained from Addgene [no. 41815]), and the sgRNAs were cloned into gRNA_GFP-T2 (a vector obtained from Addgene [no. 41820]) and a spiking vector coding for a resistance gene. After 24 to 48 h, the cells were treated with puromycin (concentration = 2 μg/ml), and resistant cells were seeded to 96-well plates using a single-cell dilution method. Growing clones were examined for the desired deletion by PCR on extracted genomic DNA (Quick Extract DNA extraction solution; Epicentre Co.), and candidates with complete loss of the WT PCR product (homozygous deletion) were screened by immunoblotting for MyoD protein.

Stable overexpression of MUNC in C2C12 cells. PCR-amplified sequence of genomic MUNC (PCR using C2C12 genomic DNA) or of spliced MUNC (PCR using cDNA from DM3 C2C12 cells) was cloned into the pLPCX vector by ligation. The constructs were linearized and introduced into the C2C12 cells (XtremeGene transfection reagent; Roche). After 24 h, pools of stably transfected cells were selected with puromycin (concentration = 2 μg/ml). Vectors coding for mutant forms of MUNC were generated similarly, using genomic DNA or DM3 cDNA as necessary.

To generate reagents for MUNC overexpression in *MYOD*$^{-/-}$ cells, the insert was cloned into the pLHCX vector by ligation. The construct was linearized and introduced into the cells (XtremeGene transfection reagent; Roche). After 48 h, pools of stably transfected cells were selected with hygromycin (concentration = 300 μg/ml).

Estimation of the proportions of spliced and unspliced MUNC in C2C12 cells transfected with genomic sequence of MUNC. To estimate the proportions of spliced and unspliced MUNC, we performed RNA-Seq from C2C12 cells stably transfected with genomic MUNC. We counted the reads overlapping three 30-base junctions made of 15 bases from each side of the exon 1-intron, exon 1-exon 2, and intron-exon 2 boundaries. The exon 1-exon 2 junction gave us an estimate of spliced MUNC, and the mean count of exon 1-intron and intron-exon 2 junctions gave an estimate for unspliced MUNC. The ratios of unspliced to spliced MUNC were 120:1 in WT C2C12 cells and 2:1 in *MYOD*$^{-/-}$ C2C12 cells.

Prediction of RNA structures. MUNC fragment structures were predicted using the Forna prediction tool (21).

TABLE 1 Primers used in this study

Primer	Sequence
qGAPDH F	GCACAGTCAAGGCCGAGAAT
qGAPDH R	GCCTTCTCCATGGTGGTGAA
qMYOD F	CATCCGCTACATCGAAGGTC
qMYOD R	GTGGAGATGCGCTCCACTAT
qMYOGENIN F	AGCGCAGGCTCAAGAAAGTGAATG
qMYOGENIN R	CTGTAGGCGCTCAATGTACTGGAT
qMYH3 F	TCCAAACCGTCTCTGCACTGTT
qMYH3 R	AGCGTACAAAGTGTGGGTGTGT
qMUNC F	AGCCTCAGGATGAGCTGTGT
qMUNC R	ATGGATGTGGGGTTCATCAT
MUNC exon1 F	TAGCCAAGGGAGCTGAAATG
MUNC exon1 R	AGTTCTCCTGCCGCCATAG
MUNC intron F	GGTTTGAAGTGCTTCCTTGG
MUNC intron R	GAGGGATGGATGTAATTGTCG
MUNC exon2 F	TATGATGAACCCCACATCCA
MUNC exon2 R	GGACGTGCTCTCTCCCATT
MUNC_HindIIIF (cloning into pLHCX)	TAAGCAAAGCTTATAGCACCTTGGAAGACTAGCCA
MUNC_HpaIR (cloning into pLHCX)	TGCTTAGTTAACTTATTCACCGAGGGACACGAAG
MUNC BglII F (cloning into pLPCX)	CTTAGATCGCAGATCTAGACTAGCCAAGGGAGCTGAA
MUNC NotI R (cloning into pLPCX)	CCGAGCTCTTGCGGCCGCTCAGTTATTCACCGAGGGACA
MUNCex1 NotI R (cloning into pLPCX)	CCGAGCTCTTGCGGCCGCACTGACCTGGAGAAGCACACAG
MUNCex2 BglII F (cloning into pLPCX)	CTTAGATCGCAGATCTTCAAATGAAAGAGCACTTATGATGA
MUNC intronic BglII F (cloning into pLPCX)	CTTAGATCGCAGATCTGTCAGTGGGCCTACAGCCTA
MUNC intronic NotI R (cloning into pLPCX)	CCGAGCTCTTGCGGCCGCACAGTGAGGGATGGATGTAATTGT
sgMYOD1	AGCTTCTATCGCCGCCACTCCGG
sgMYOD2	TGTAGCGGATGGCGTTGCGCAGG
MYODcrisprKO_F	CGAAGCTATGGAGCTTCTATCGCCGCCA
MYODcrisprKO_R	CCTTACCATGCCATCAGAGCAGTTGGAG
sgMUNC_1	CACCTTGGAAGACTAGCCAAGGG
sgMUNC_2	GCATACCATGGATAGGAGTATGG
MUNCcrisprKO_F	CTTGAGTTGGGAAAGGAAAGTCTAGGG
MUNCcrisprKO_R	GTCTCAGATCTCAACTCCAAAGTCATTTTT
Tnnc1F	GAAGGACGACAGCAAAGGGA
Tnnc1R	AGCCATCAGCGTTTTTGTCA
Tmem8cF	GCTGGAGAAGCAAAGAAGTGG
Tmem8cR	CTACAACTGTCCCCATGGACC
Ablim3F	CTGGCCAAGAGGTGATGAGT
Ablim3R	GCTCGTGTTCATGGTGATGC
MylpfF	ACCACGGTATGTTAAGGGCTG
MylpfR	TCTTAGATCTCCTGGGGGCAA
MUNC probe F	TGCCCTCCAAATGGATCACC
MUNC probe R	CAGCAGTAAGCGCAACCAAG
MyoD1_pCW_F	TGGAGAATTGGCTAGCGCCGCCATGGAGCTTCTATCGCCGCC
MyoD1_pCW_R	CCCCAACCCCGGATCCTCAAAGCACCTGATAAATCG
sg_miR1-1—1	TGCACAAGAACAGGACTCCGAGG
sg_miR1-1—2	GCATGGGCCACCCCTCAGTCTGG

Transient overexpression of *MYOD* in C2C12 cells. Cells were seeded on 6-well plates and after 12 h were transfected with vector coding for *MYOD*. The medium was changed 12 h posttransfection to differentiation medium, and cells were harvested 2 days later.

Stable overexpression of inducible *MYOD* in WT, *MUNC*$^{-/-}$, and *MYOD*$^{-/-}$ C2C12 cells. PCR-amplified sequence of the *MYOD* ORF (PCR using C2C12 cDNA) was cloned by ligation into the pCW-Cas9 (Addgene; no. 50661) vector upon Cas9 removal by enzymatic digestion with BamHI and NheI. The vector was packed in the virus using psPAX2 (Addgene; no. 12260) and pMD2.G (Addgene; no. 12259) in 293T cells. WT, *MUNC*$^{-/-}$, and *MYOD*$^{-/-}$ C2C12 cells were transduced with the filtered supernatant containing virus. After 24 h, the cells were treated with puromycin ($C = 2$ μg/ml).

MYOD expression was induced in *MUNC*$^{-/-}$ and *MYOD*$^{-/-}$ C2C12 cells before differentiation using doxycycline (concentration = 1 μg/ml). The samples were collected under proliferation (GM) and differentiation (DM1, DM3, and DM5) conditions.

RNA analysis by qRT-PCR. RNA was isolated by TRIzol extraction or using an RNeasy minikit (Qiagen), and RNA samples were treated with RQ1 RNase-free DNase (Promega Co.) to eliminate potential DNA contamination of samples. cDNA synthesis was performed using a Superscript III RT cDNA synthesis kit (Life Technologies Co.) with random-hexamer and oligo(dT) priming. After cDNA synthesis, quantitative PCR (qPCR) was performed with Applied Biosystems 7500 real-time PCR systems using Power SYBR green master mix (ThermoFisher Scientific) or a SensiFast SYBR Hi-Rox kit (Bioline). All the primers used in this study are listed in Table 1.

Western blotting. Cells were lysed in IPH buffer (50 mM Tris-Cl, 0.5% NP-40, 50 mM EDTA), run on a 10% polyacrylamide SDS-PAGE gel, and transferred to nitrocellulose membranes. The membranes were blocked for 30 min in 5% milk containing phosphate-buffered saline with Tween 20 (PBS-T) and incubated overnight with primary antibody in 1% milk. Secondary-antibody incubation was carried out for 1 h after washing and at 1:4,000 dilution before washing and incubation with Millipore Immobilon

horseradish peroxidase (HRP) substrate. Antibodies were used as follows: MyoD1 (sc-12732; Santa Cruz Co.), MHC (MF-20; Developmental Studies Hybridoma Bank, University of Iowa), Ablim3 (sc-398575; Santa Cruz Co.), Mylpf (16052-1-AP; Proteintech Co.), and Tnnc1 (13504-1-AP; Proteintech Co.).

Southern blotting. Ten micrograms of genomic DNA was digested with a restriction enzyme and electrophoresed in a 0.8% agarose gel. The DNA was transferred to a Nitran SuperCharge membrane (Schleicher & Schuell) using alkaline denaturing conditions. The membrane was hybridized with a DNA probe labeled with a random-primer DNA-labeling kit (TaKaRa) using [^{32}P]dCTP. The probe was amplified from genomic DNA with MUNC probe forward and MUNC probe reverse primers (listed in Table 1).

Immunofluorescence assay. Cells were plated on glass coverslips and collected in growth medium or after 3 days of differentiation. The coverslips were fixed with 4% formaldehyde in PBS for 10 min, permeabilized in 0.5% Triton X-100 in PBS, and blocked in 5% goat serum. The coverslips were incubated at room temperature with primary antibody for 1 h and Alexa Fluor 488- or 549-conjugated secondary antibody for 1 h, with three PBS washes following each antibody incubation. The coverslips were then mounted with Vectashield mounting solution (Vector Laboratories). The antibodies used were anti-MyoD C-20 antibody (Santa Cruz Laboratories) and anti-myosin heavy chain M4276 antibody (Sigma). The antibodies were diluted 1:200 in 5% goat serum containing PBS.

Microscopy. Images were captured using a Nikon Microphot SA upright microscope equipped with a Nikon NFX35 camera using SPOT imaging software (Diagnostic Instruments Inc.) and a Nikon PlanApo 60× oil objective lens. Fluorescence images were acquired on the same day using the same exposure times, gamma, and gain between samples. Images were enhanced for brightness and contrast to the same extent within Adobe Photoshop software.

RNA-Seq library preparation. RNA samples were isolated from proliferating or differentiating cells using an RNeasy minikit (Qiagen Co.). One microgram of RNA was enriched for poly(A)-tailed mRNA molecules using a NEBNext Poly(A) mRNA Magnetic Isolation Module, and RNA-Seq libraries were made using NEBNext Ultra Directional RNA library prep kit for Illumina (NEB Co.) according to the manufacturer's protocol. Pooled libraries were sequenced using a paired-end protocol on the Illumina platform, using a NextSeq 500 instrument in the Biomolecular Analysis Facility, University of Virginia School of Medicine.

RNA-Seq analysis. We obtained ≥40 million paired-end 75-bp-long reads for WT and *MYOD* knockout (*MYOD*$^{-/-}$) conditions. The WT control cell line, the WT cell line overexpressing MUNC, the *MYOD*$^{-/-}$ cell line, and the *MYOD*$^{-/-}$ cell line overexpressing MUNC were grown in GM conditions and harvested at ~80% confluence. To achieve differentiated samples (DM) at ~90% confluence of cells, medium was changed to differentiation medium, and cells were harvested after 3 days. Paired-end reads were obtained from the two biological replicates with EV and MUNC overexpression in both GM and DM in WT and *MYOD*$^{-/-}$ C2C12 cell lines. Transcripts for mm10 RefSeq genes were downloaded from the UCSC table browser (http://genome.ucsc.edu). We used the default settings of Kallisto (42) to build an index for the downloaded 35,818 transcript sequences and then quantified the abundance of each transcript from the paired-end reads (42). We used the DESeq2 package in R for differential expression analysis of the quantified data obtained from Kallisto (43). A *P* value (obtained by DESeq2) cutoff of 0.05 was used to define differentially expressed genes. Gene Trail (44) and GSEA (45) were used for functional gene ontology term enrichment analysis and gene set enrichment analysis, respectively.

Accession number(s). All RNA-Seq library data files are available under GEO accession number GSE99258.

ACKNOWLEDGMENTS

We thank all the members of the Dutta laboratory for helpful discussions, especially Etsuko Shibata for technical guidance regarding the CRISPR-Cas9 method.

This work was supported by a grant from the NIH (NIAMS), R01 AR067712, to A.D. M.A.C. was partly supported by a Wagner Fellowship from the University of Virginia during the study. M.K. was partly supported by a DOD award (PC151085).

M.A.C. performed structure-function studies. R.K.P. analyzed the structure predictions. M.A.C. designed CRISPR-Cas9 for MYOD and MUNC KO and obtained the knockout cells with help from R.K.P. R.K.P. designed CRISPR-Cas9 for miR-1a-1 KO and obtained the knockout cells with help from E.S. M.A.C. performed immunofluorescence experiments. M.A.C., R.K.P., and E.S. performed differentiation assays and qPCR analyses. M.A.C. and R.K.P. performed Western blot analyses. Y.S. performed Southern blotting experiments. M.A.C. designed MUNC stable-overexpression studies. R.K.P. designed MYOD stable-overexpression studies. M.A.C. performed the RNA-Seq experiment with help from R.K.P. M.K. performed analysis related to Fig. 7 and 9. M.A.C. confirmed RNA-Seq results. M.A.C. wrote the manuscript with help from A.D., M.K., and R.K.P.

REFERENCES

1. Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R. 2010. Long noncoding RNAs with enhancer-like function in human cells. Cell 143: 46–58. https://doi.org/10.1016/j.cell.2010.09.001.

2. Engreitz AJM, Haines JE, Munson G, Chen J, Elizabeth M, Kane M, Mcdonel PE, Guttman M, Lander ES. 2016. Neighborhood regulation by lncRNA promoters, transcription, and splicing. Nature 539:1–15. https://doi.org/10.1038/nature20149.
3. Braun T, Gautel M. 2011. Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. Nat Rev Mol Cell Biol 12:349–361. https://doi.org/10.1038/nrm3118.
4. Yokoyama S, Asahara H. 2011. The myogenic transcriptional network. Cell Mol Life Sci 68:1843–1849. https://doi.org/10.1007/s00018-011-0629-2.
5. Berkes CA, Tapscott SJ. 2005. MyoD and the transcriptional control of myogenesis. Semin Cell Dev Biol 16:585–596. https://doi.org/10.1016/j.semcdb.2005.07.006.
6. Tapscott SJ, Lassar AB, Weintraub H. 1992. A novel myoblast enhancer element mediates MyoD transcription. Mol Cell Biol 12:4994–5003. https://doi.org/10.1128/MCB.12.11.4994.
7. Asakura A, Lyons GE, Tapscott SJ. 1995. The regulation of MyoD gene-expression; conserved elements mediate expression in embryonic axial muscle. Dev Biol 171:386–398. https://doi.org/10.1006/dbio.1995.1290.
8. Goldhamer DJ, Faerman A, Shani M, Emerson CP. 1992. Regulatory elements that control the lineage-specific expression of myoD. Science 256:538–542. https://doi.org/10.1126/science.1315077.
9. Chen JCJ, Ramachandran R, Goldhamer DJ. 2002. Essential and redundant functions of the MyoD distal regulatory region revealed by targeted mutagenesis. Dev Biol 245:213–223. https://doi.org/10.1006/dbio.2002.0638.
10. L'honore A, Lamb NJ, Vandromme M, Turowski P, Carnac G, Fernandez A. 2003. MyoD distal regulatory region contains an SRF binding CArG element required for MyoD expression in skeletal myoblasts and during muscle regeneration. Mol Biol Cell 14:2151–2162. https://doi.org/10.1091/mbc.e02-07-0451.
11. Chen JCJ, Love CM, Goldhamer DJ. 2001. Two upstream enhancers collaborate to regulate the spatial patterning and timing of MyoD transcription during mouse development. Dev Dyn 221:274–288. https://doi.org/10.1002/dvdy.1138.
12. Mousavi K, Zare H, Dell'Orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager G, Sartorelli V. 2013. ERNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell 51:606–617. https://doi.org/10.1016/j.molcel.2013.07.022.
13. Mueller AC, Cichewicz MA, Dey BK, Layer R, Reon BJ, Gagan JR, Dutta A. 2015. MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis. Mol Cell Biol 35:498–513. https://doi.org/10.1128/MCB.01079-14.
14. Ballarino M, Morlando M, Fatica A, Bozzoni I. 2016. Non-coding RNAs in muscle differentiation and musculoskeletal disease. J Clin Invest 126:2021–2030. https://doi.org/10.1172/JCI84419.
15. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, Yamanaka I, Kiyosawa H, Yagi K, Tomaru Y, Hasegawa Y, Nogami A, Schonbach C, Gojobori T, Baldarelli R, Hill DP, Bult C, Hume D, Quackenbush J, Schriml LM, Kanapin A, Matsuda H, Batalov S, Beisel KW, Blake J, Bradt D, Brusic V, Chothia C, Corbani LE, Cousins S, Dalla E, Dragani T, Fletcher CF, Forrest A, Frazer KS, Gaasterland T, Gariboldi M, Gissi C, Godzik A, Gough J, Grimmond S, Gustincich S, Hirokawa N, Jackson IJ, Jarvis ED, Kanai A, Kawaji H, Kawasawa Y, Kedzierski RM, King BL, Konagaya A, Kurochkin IV, Lee Y, Lenhard B, Lyons P, Maglott DR, Maltais L, Marchionni L, McKenzie L, Miki H, Nagashima T, Numata K, Okido T, Pavan WJ, Pertea G, Pesole G, Petrovsky N, Pillai R, Pontius JU, Qi D, Ramachandran S, Ravasi T, Reed JC, Reed DJ, Reid J, Ring BZ, Ringwald M, Sandelin A, Schneider C, Semple C, Setou M, Shimada K, Sultana R, Takenaka Y, Taylor MS, Teasdale RD, Tomita M, Verardo R, Wagner L, Wahlestedt C, Wang Y, Watanabe Y, Wells C, Wilming LG, Wynshaw-Boris A, Yanagisawa M, Yang I, Yang L, Yuan Z, Zavolan M, Zhu Y, Zimmer A, Carninci P, Hayatsu N, Hirozane-Kishikawa T, Konno H, Nakamura M, Sakazume N, Sato K, Shiraki T, Waki K, Kawai J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashizume W, Imotani K, Ishii Y, Itoh M, Kagawa I, Miyazaki A, Sakai K, Sasaki D, Shibata K, Shinagawa A, Yasunishi A, Yoshino M, Waterston R, Lander ES, Rogers J, Birney E, Hayashizaki Y, FANTOM Consortium, RIKEN Genome Exploration Research Group Phase I and II Team. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420:563–573. https://doi.org/10.1038/nature01266.
16. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515. https://doi.org/10.1038/nbt.1621.
17. Neguembor MV, Jothi M, Gabellini D. 2014. Long noncoding RNAs, emerging players in muscle differentiation and disease. Skelet Muscle 4:8. https://doi.org/10.1186/2044-5040-4-8.
18. Colley SM, Leedman PJ. 2011. Steroid receptor RNA activator—a nuclear receptor coregulator with multiple partners: insights and challenges. Biochimie 93:1966–1972. https://doi.org/10.1016/j.biochi.2011.07.004.
19. Wang L, Zhao Y, Bao X, Zhu X, Kwok YK-Y, Sun K, Chen X, Huang Y, Jauch R, Esteban MA, Sun H, Wang H. 2015. LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. Cell Res 25:335–350. https://doi.org/10.1038/cr.2015.21.
20. Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, Gabellini D. 2012. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell 149:819–831. https://doi.org/10.1016/j.cell.2012.03.035.
21. Kerpedjiev P, Hammer S, Hofacker IL. 2015. Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. Bioinformatics 31:3377–3379. https://doi.org/10.1093/bioinformatics/btv372.
22. Yu X, Zhang Y, Li T, Ma Z, Jia H, Chen Q, Zhao Y, Zhai L, Zhong R, Li C, Zou X, Meng J, Chen AK, Puri PL, Chen M, Zhu D. 2017. Long non-coding RNA Linc-RAM enhances myogenic differentiation by interacting with MyoD. Nat Commun 8:14016. https://doi.org/10.1038/ncomms14016.
23. Megeney LA, Kablar B, Garrett K, Anderson JE, Rudnicki MA. 1996. MyoD is required for myogenic stem cell function in adult skeletal muscle. Genes Dev 10:1173–1183. https://doi.org/10.1101/gad.10.10.1173.
24. Rajan S, Dang HCP, Djambazian H, Zuzan H, Fedyshyn Y, Ketela T, Moffat J, Hudson TJ, Sladek R. 2012. Analysis of early C2C12 myogenesis identifies stably and differentially expressed transcriptional regulators whose knock-down inhibits myoblast differentiation. Physiol Genomics 44:183–197. https://doi.org/10.1152/physiolgenomics.00093.2011.
25. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8:2281–2308. https://doi.org/10.1038/nprot.2013.143.
26. Conerly ML, Yao Z, Zhong JW, Groudine M, Tapscott SJ. 2016. Distinct activities of Myf5 and MyoD indicate separate roles in skeletal muscle lineage specification and differentiation. Dev Cell 36:375–385. https://doi.org/10.1016/j.devcel.2016.01.021.
27. Wang Y, Szczesna-Cordary D, Craig R, Diaz-Perez Z, Guzman G, Miller T, Potter JD. 2007. Fast skeletal muscle regulatory light chain is required for fast and slow skeletal muscle development. FASEB J 21:2205–2214. https://doi.org/10.1096/fj.06-7538com.
28. Matsuda M, Yamashita JK, Tsukita S, Furuse M. 2010. AbLIM3 is a novel component of adherens junctions with actin-binding activity. Eur J Cell Biol 89:807–816. https://doi.org/10.1016/j.ejcb.2010.07.009.
29. Perry SV. 1985. Properties of the muscle proteins—a comparative approach. J Exp Biol 115:31–42.
30. Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Vrielink JAFO, Elkon R, Melo SA, Léveillé N, Kalluri R, de Laat W, Agami R. 2013. ERNAs are required for p53-dependent enhancer activity and gene transcription. Mol Cell 49:524–535. https://doi.org/10.1016/j.molcel.2012.11.021.
31. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen Y, Merkurjev D, Zhang J, Ohgi K, Song X, Kim H, Glass CK, Rosenfeld MG. 2013. Functional importance of eRNAs for estrogen-dependent transcriptional activation events. Nature 498:516–520. https://doi.org/10.1038/nature12210.
32. Kim TK, Hemberg M, Gray JM. 2015. Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. Cold Spring Harb Perspect Biol 7:a018622. https://doi.org/10.1101/cshperspect.a018622.
33. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson EN. 2015. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 160:595–606. https://doi.org/10.1016/j.cell.2015.01.009.
34. Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F, Reese AL, McAnally JR, Chen X, Kavalali ET, Cannon SC, Houser SR, Bassel-Duby R, Olson EN. 2016. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science 351:271–275. https://doi.org/10.1126/science.aad4076.
35. Anderson DM, Makarewich CA, Anderson KM, Shelton JM, Bezprozvannaya S, Bassel-Duby R, Olson EN. 2016. Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. Sci Signal 9:ra119. https://doi.org/10.1126/scisignal.aaj1460.
36. Mal A, Harter ML. 2003. MyoD is functionally linked to the silencing of a

muscle-specific regulatory gene prior to skeletal myogenesis. Proc Natl Acad Sci U S A 100:1735–1739. https://doi.org/10.1073/pnas.0437843100.
37. Bengal E, Ransone L, Scharfmann R, Dwarki VJ, Tapscott SJ, Weintraub H, Verma IM. 1992. Functional antagonism between c-Jun and MyoD proteins: A direct physical association. Cell 68:507–519. https://doi.org/10.1016/0092-8674(92)90187-H.
38. Battistelli C, Busanello A, Maione R. 2014. Functional interplay between MyoD and CTCF in regulating long-range chromatin interactions during differentiation. J Cell Sci 127:3757–3767. https://doi.org/10.1242/jcs.149427.
39. Rudnicki MA, Schnegelsberg PNJ, Stead RH, Braun T, Arnold HH, Jaenisch R. 1993. MyoD or Myf-5 is required for the formation of skeletal muscle. Cell 75:1351–1359. https://doi.org/10.1016/0092-8674(93)90621-V.
40. Kablar B, Krastel K, Ying C, Asakura A, Tapscott SJ, Rudnicki M. 1997. MyoD and Myf-5 differentially regulate the development of limb versus trunk skeletal muscle. Development 124:4729–4738.
41. Hannon K, Smith CK, Bales KR, Santerre RF. 1992. Temporal and quantitative analysis of myogenic regulatory and growth factor gene expression in the developing mouse embryo. Dev Biol 151:137–144. https://doi.org/10.1016/0012-1606(92)90221-2.
42. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–527. https://doi.org/10.1038/nbt.3519.
43. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8.
44. Keller A, Backes C, Al-Awadhi M, Gerasch A, Küntzer J, Kohlbacher O, Kaufmann M, Lenhof H-P. 2008. GeneTrailExpress: a web-based pipeline for the statistical evaluation of microarray experiments. BMC Bioinformatics 9:552. https://doi.org/10.1186/1471-2105-9-552.
45. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette M, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545–15550. https://doi.org/10.1073/pnas.0506580102.
46. Lai F, Orom U, Cesaroni M, Beringer M, Taatjes DJ, Blobel G, Shiekhattar R. 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494:497–501. https://doi.org/10.1038/nature11884.

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

LINC00152 Promotes Invasion Through a 3'-hairpin Structure and Associates with Prognosis in Glioblastoma

Brian J Reon [*], Bruno Takao Real Karia [*], Manjari Kiran and Anindya Dutta [1]

Department of Biochemistry and Molecular Genetics, University of Virginia,

Charlottesville, VA, USA.

[*]These authors contributed equally

[1]: correspondence to ad8q@virginia.edu

Running title: **lncRNA LINC00152 Promotes Invasion in GBM**

The authors declare no potential conflicts of interest.

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Abstract

Long non-coding RNAs (lncRNAs) are increasingly implicated in oncogenesis. Here, it is determined that *LINC00152*/*CYTOR* is upregulated in glioblastoma multiforme (GBM) and aggressive wild-type IDH1/2 grade II/III gliomas and upregulation associates with poor patient outcomes. *LINC00152* is similarly upregulated in over 10 other cancer types and associates with a poor prognosis in 7 other cancer types. Inhibition of the mostly cytoplasmic *LINC00152* decreases, and overexpression increases cellular invasion. *LINC00152* knockdown alters the transcription of genes important to epithelial-to-mesenchymal transition (EMT). PARIS and Ribo-seq data, together with secondary structure prediction, identified a protein bound 121bp stem-loop structure at the 3' end of *LINC00152* whose overexpression is sufficient to increase invasion of GBM cells. Point mutations in the stem-loop suggest that stem formation in the hairpin is essential for *LINC00152* function. *LINC00152* has a nearly identical homolog, *MIR4435-2HG*, which encodes a near identical hairpin, is equally expressed in low-grade glioma (LGG) and GBM, predicts poor patient survival in these tumors and is also reduced by *LINC00152* knockdown. Together, these data reveal that *LINC00152* and its homolog *MIR4435-2HG* associate with aggressive tumors and promote cellular invasion through a mechanism that requires the structural integrity of a hairpin structure.

Implications: Frequent upregulation of the lncRNA, *LINC00152*, in glioblastoma and other tumor types combined with its prognostic potential and ability to promote invasion suggests *LINC00152* as a potential biomarker and therapeutic target.

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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 [1]. 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 [2]. These efforts have identified commonly altered signaling pathways in GBMs, including mutations in EGFR, p53 and mTOR signaling [3,4]. 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 [5,6]. 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 surgical resection, radiotherapy and Temozolomide, still leaves patients with a 5-year survival rate of roughly 10% [7].

High throughput sequencing revealed that a majority of the human genome, 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 [8]. 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 subclasses based on chromosomal position relative to other genes, enhancers or other genomic regulatory elements. LncRNAs have been shown to play many different functional roles in the cell, in part through regulation of transcription, mRNA stability and mRNA translational efficiency [9,10].

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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 prostate cancer [11-13].

LINC00152 is a lncRNA that was first identified as being hypomethylated during hepatocellular carcinoma tumorigenesis [14]. It is also dysregulated in gastric cancer and esophageal squamous cell carcinoma [15,16]. 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 competing endogenous RNA (ceRNA) through titrating microRNAs [5,6,17-20]. 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 and with tumor cell invasion and begin to functionally characterize *LINC00152* structurally. Furthermore, we find that *LINC00152* is overexpressed in 10 other tumor types compared to matched normal tissue and high *LINC00152* expression is associated with a poor prognosis in 7 of these tumors.

Materials and Methods

Cell culture, knockdown and overexpression of *LINC00152*

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

U87 cells were maintained in MEM supplemented with 1% non-essential amino acids solution (cat # 11140-050, Gibco), 1mM sodium pyruvate (cat # 11360070, Gibco), 0.15% sodium bicarbonate (cat # 25080094, Gibco), 10% FBS and 1% P/S.

For knockdown, U87 cells were transfected during two rounds of transfection. First, cells were reverse transfected with 40 ηM of si*LINC00152*_II (5'-UGACACACUUGAUCGAAUA-3'), si*LINC00152*_III (5'-CCGGAAUGCAGCUGAAAGA-3') or a nonspecific siGL2 control siRNA (5'-CGUACGCGGAAUACUUCGA-3') and 9 µL of Lipofectamine RNAiMAX transfection reagent (Thermo Fisher). 24 hours later, a second round of transfection was performed using the same quantities of reagents. 24 hours after the final transfection, cells were harvested and used for subsequent analysis.

500ηg of *LINC00152* or *LINC00152* mutants pCDNA3-flag vectors were transfected into U87 cells using 2µL of Lipofectamine 2000 (Thermo Fisher). Cells were harvested after 48 hours for downstream analysis.

RNA isolation, cDNA synthesis, qPCR and Western blotting

Total RNA and nuclear/cytoplasmic RNAs were extracted using TRIzol total RNA isolation reagent (Thermo Fisher), Protein and RNA Isolation System (ThermoFisher), respectively. RNA samples were treated with RQ1 RNase-Free DNase (Promega) according to according to manufacturer's instructions. cDNA was produced from 1µg RNA using Superscript III kit (Thermo Fisher) according to manufacturer's instructions. qPCR and Western blotting were performed according to standard protocols.

***LINC00152* subcellular fractionation and *in situ* hybridization**

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

LINC00152 subcellular fractionation was performed using Protein and RNA Isolation System (ThermoFisher) according to manufacturer instructions.

For *in situ* hybridization $3x10^5$ U87 and U251 cells were plated on the top of a cover glass in a 6-well plate. In the next day, cells were washed once with PBS and fixed for 10min with 2% paraformaldehyde. Then, cells were washed 3 times with PBS and incubated with 1mL of permeabilization buffer (1× PBS/0.5% Triton X-100) for 10min at 4ºC. Cells were again washed 3 times with PBS. Next, cells were blocked with 1mL of prehybridization buffer (3% BSA in 4× SSC) for 20min at 55ºC. 10ηg of *LINC00152* or negative control probe were added to 2mL of hybridization buffer (10% dextran sulfate in 4× SSC) and cells were incubated overnight at 55ºC. On the next day, cells were washed 3 times for 5min using washing buffer I (4× SSC, 0.1% Tween-20), 3 times for 5min using washing buffer II (2× SSC), and 3 times for 5min using washing buffer III (1× SSC). Subsequently, cells were blocked for 15min at room temperature using 2mL of blocking solution (4% BSA/1× PBS). Next, cells were incubated with 300μL of antibody solution [2% of BSA/1× PBS and digoxigenin (1:250)] for 1h. Cells were washed with 0.1% Tween-20/PBS 3 times and with alkaline Tris buffer for 5min at room temperature. Finally, the signal was developed by adding 400μL of BCP/NBT solution until the signal was visible.

MTT and matrigel invasion assays

For measuring cell growth, 1,000 cells were plated in quadruplets in 96 well plates and cell growth was measured using standard MTT reagent (Promega). To measure invasion, $2x10^5$ 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%

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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.

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 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 for the remaining tumor types were retrieved from OncoLnc (www.oncolnc.org) on 12/2016 [22-24]. The expression threshold used to separate patients are outlined in the main text. Kaplan Meier plots, hazard ratios and p-values, based on these separations were generated using the 'survminer' package for R.

RNA-seq analysis

U87 cells were treated with a combination of the two 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 75 bp paired-end reads NextSeq500 instrument, in the Biomolecular Analysis Facility, University of Virginia School of Medicine. Sequencing reads were aligned to the hg38 reference genome using HISAT [25]. Gene abundances and identification of

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

differentially expressed genes were performed using HTSeq and DESeq2 [26,27]. An adjusted P-value (obtained by DESeq2) cut-off of 0.05 and Log 2-Fold change of 2 was used to define differentially expressed genes. GSEA analysis was performed on preranked gene list based on fold change (si*LINC00152*/siGL2) against 50 hallmark gene sets [28]. For plotting enrichment score obtained by GSEA analysis (as shown in Fig 4B), we have included only the genes which are either induced or repressed 1.5-fold suupon siLINC00152. The raw and processed data were deposited in Gene Expression Omnibus (GEO) under accession number GSE111652.

***LINC00152* structure predictions**

Secondary structure predictions of *LINC00152* were determined using mfold [29]. 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.* [30,31]. Aligned reads were then processed to identify gapped mapping to *LINC00152* and visualized with IGV [32]. We used ribosome profiling data from Gonzalez *et. al.* and aligned reads to the hg19 genome using HISAT2 [33]. 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 [32]. The predicted secondary structure elements and protein bound region were then compared to the *in silico* secondary structure predictions.

Results

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

LINC00152 is a lncRNA overexpressed 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 1A and Sup Fig 1A). We have validated the upregulation of *LINC00152* in an independent set of GBM patients compared to normal FFPE brain tissue [21].

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 1B). Even though *LINC00152* is not upregulated in LGGs as a whole, the IDHwt LGG subtype expresses 4 times as much *LINC00152* as normal brains ($p < 0.00001$) (Fig 1B). This is interesting, because IDHwt LGGs are far more aggressive than the other LGG subtypes and display clinical properties similar to GBMs [34].

LINC00152 expression predicts survival in GBMs and LGGs

Since *LINC00152* is upregulated in brain tumors compared to normal brain tissue, we next elucidated the association of *LINC00152* association with survival of GBM and LGG patients. To do this, we assessed the survival difference of patients expressing high (top 33% highest expressing LINC00152 cohort) and low level of *LINC00152* expression from the TCGA for both GBM and LGG. In GBMs, patients who had high expression of *LINC00152* had a poor prognosis ($p = 0.02$) compared to the patients expressing low level of *LINC00152*, with a median survival of 11.9 and 15.4 months,

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

respectively (Fig 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 of 98.2 months ($p < 0.0001$) (Fig 2B). These results demonstrate that not only is *LINC00152* overexpressed in gliomas, but that this overexpression is associated with poor patient outcome.

LINC00152 **in other cancers**

It was intriguing to examine *LINC00152* expression in other cancers compared to their respective normal tissues. 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 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 (Fig 1C-L).

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, lung adenocarcinoma, renal clear cell carcinoma and

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

hepatocellular carcinoma (Fig 2C-F). The poor 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 poor outcome was statistically significant (p = 0.015) when we compared patients in the top third and bottom third based on *LINC00152* expression (Sup Fig 1C).

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 when we compare the tumors in the top third and bottom third (Sup Fig 1D), and acute myeloid leukemia, with the top quartile and bottom quartile for *LINC00152* expression (Sup Fig 1E). 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 unfavorable gene whose expression is associated with a poor patient outcome.

***LINC00152* expression controls GBM cell invasion**

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Subcellular fractionation (Fig 3A and B) and *in-situ* hybridization (Fig 3C) revealed that *LINC00152* is primarily localized in the cytoplasm of U87 cells. 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) [35,36]. We knocked down *LINC00152* expression using two separate siRNAs or overexpressed the lncRNA and found that *LINC00152* knockdown or overexpression did not affect cell proliferation for a period of 10 days (Sup Fig 2B and D). 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 3D and E). Conversely, overexpression of *LINC00152* led to an increase of over 2-fold in the number of invaded cells (Fig 3F and G). These findings suggest that *LINC00152* knockdown decreases invasion of GBM cells, while upregulation in GBMs promotes the invasive phenotype that is commonly seen in patient tumors.

***LINC00152* knockdown decreases expression of pro-invasive genes.**

In order to better understand how *LINC00152* affects cellular invasion we performed RNA-seq on U87 following knockdown of *LINC00152* using a combination of two different siRNAs. Knockdown of *LINC00152* leads to large changes in gene expression, with 259 genes significantly up-regulated and 295 down-regulated at least 2-fold (Fig 4A). Thus, to determine the most significant molecular pathways regulated by *LINC00152*, we performed GSEA (gene set enrichment analysis), a method that can

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

identify pathway enrichment from fold change based pre-ranked gene list from RNA-seq [28]. This analysis showed a significant enrichment of up-regulated genes upon si*LINC00152* involved in Epithelial to Mesenchymal transition (EMT) (Fig 4B). Among the differentially expressed genes involved in EMT, the changes were validated by qPCR on 12 out of 13 genes after siLIN00152 treatment (Sup Table 1). More interestingly, six of the genes that were downregulated by *LINC00152* knockdown were conversely upregulated by overexpression of the lncRNA: *TPM2* (Tropomyosin 2), *PTX3* (Pentraxin 3), *IGFBP4* (Insulin growth factor binding protein 4), *TGM2* (Transglutaminase 2), *SPP1* (Secreted phosphoprotein 1) and *LUM* (Lumican)] (Sup Table 1). Moreover, overexpression of the siRNA-resistant M8 was sufficient to upregulate these genes even after knockdown of endogenous *LINC00152* (Sup Fig 3). These results indicate that *LINC00152* may induce U87 cells invasion by regulating the expression of at least these six genes.

LINC00152 **is not involved in sponging of miRNAs**

Several previous studies have suggested that *LINC00152* acts as a microRNA sponge by titrating different microRNAs (miR-376c-3p, miR-4775, miR-4767, miR-138-5p, miR-103 and miR-205) in different types of tumors, including GBMs [5,6,17-20]. However, suggestions that an lncRNA acts as a microRNA sponge are sometimes questioned because the abundance of the lncRNA is often far less than that of the targets of the microRNAs and of the microRNAs themselves. If *LINC00152* acts as a miRNA sponge in U87 cells we would expect that the targets of these microRNAs would be repressed upon knockdown of the lncRNA and the subsequent release of the microRNAs from interaction with the lncRNA. However, we find that there is a statistically significant up-

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

regulation of the targets of these six microRNAs compared with non-targets when *LINC00152* is knocked down ruling out the possibility of *LINC00152* acting as a ceRNA for these miRNAs (Fig 4C).

Secondary structure components of *LINC00152*

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) [30]. PARIS is based on reversibly crosslinking RNA duplexes (stems of stem-loops) and gentle digestion with a single-strand RNase, 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 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 the PARIS libraries (Sup Fig 4C). These reads are positioned from position 285 to 373 of the 496 nt long *LINC00152*, with a small 2 base gap starting at position 342 (Fig 5A). Sequence analysis of this region revealed some complementarity, suggesting that this region might in fact form a stem-loop structure (Fig 5A).

To get a better understanding of overall *LINC00152* secondary structure, we used publicly available RNA secondary structure prediction tool, mfold, to identify secondary structure predictions for *LINC00152* that are consistent with a stem-loop being present

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

from 285-373 [29]. 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 285 to 373 (Fig 5A and Sup Fig 4A and 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 [37]. This information has also been used to ascertain that some lncRNAs are associated with ribosomes, but not translating ribosomes [38]. Recently it was determined that the polysomes isolated for Ribo-seq are contaminated with other ribonucleoprotein (RBP) complexes. As a result RNA footprints from RBPs that are not ribosome proteins can be detected in Ribo-seq data [33]. To determine if we could identify RBP-RNA footprints from *LINC00152* we analyzed publicly available Ribo-seq data from normal brain samples [39]. In two out of the three normal brain Ribo-seq samples we detected a RBP footprint at positions 303-330 of *LINC00152*. In addition, in one of the samples there was an RBP footprint from 354-382 (Fig 5A and Sup Fig 4D). These two footprinted areas are located on opposite strands of the same stem-loop that was detected by PARIS, providing additional evidence of the existence of this stem-loop and suggesting that this stem is bound by a protein in an RBP (Fig 5A).

***LINC00152* stem-loop, M8, is sufficient to promote cell invasion.**

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

In order to determine whether this newly identified, potentially protein bound, stem-loop plays a role in *LINC00152* function, we created a series of *LINC00152* deletion mutants (Fig 5B and Sup Fig 5A). The sites of the deletions were chosen based on PARIS and Ribo-seq analysis as well as two *in silico* predicted structures of *LINC00152* (Fig 5A and Sup Fig 4C and D). We assessed whether independent overexpression of the mutants was able to stimulate U87 cell invasion. Overexpression of M2 (which removed the minimal amount of the protein bound stem-loop, nucleotides 280-401) or M3 (which removed the stem-loop and the remaining 3' end) led to a decreased cell invasion significantly compared to full-length *LINC00152* ($p < 0.05$) (Fig 5D). On the other hand, the mutant M4 (which removed the 3' end but preserved the stem-loop) or M7 (which removed the extreme 3' end, and also preserved the stem-loop) increased U87 cell invasion. Other deletion mutants that removed regions of *LINC00152* 5' to the stem loop (M5 or M6) stimulated cellular invasion to a similar extent as full-length *LINC00152*. Finally, overexpression of M8, containing only the protein bound stem-loop (nucleotides 280-401) was sufficient to stimulate invasion of U87 cells (Fig 5D). These results suggest that M8 stem-loop is necessary and sufficient for stimulation of cell invasion.

Consistent with this conclusion, overexpression of the stem-loop also induced the six genes involved in EMT to the same extent as the full length *LINC00152* (Sup Table 1). In addition, knockdown of *LINC00152* by si*LINC00152*_II (a siRNA that targets a region on *LINC00152* outside of M8) decreased cell invasion while the siRNA-resistant M8 was sufficient to rescue cell invasion (Fig 5E).

The M8 stem-loop structure is important for stimulating invasion.

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

We next tested with point mutations whether the ability to stimulate invasion of U87 cells depends on the *LINC00152* stem-loop structure. Two mutants on opposite side of the stem disrupt the stem-loop (mutA: changes bases 333-336 and mutB: changes bases 349-352) (Fig 6D). Neither mutA nor mutB stimulated the invasion of U87 cells as well as full length *LINC00152* or M8 (Fig 6A). In contrast, when the two mutations were combined in mutAB, the stem-loop structure was reconstructed and this promoted invasion to the same extent as full length *LINC00152* or M8 (Fig 6A). Therefore, we can conclude that the stem-loop structure itself is essential for *LINC00152* to stimulate cellular invasion.

MIR4435-2HG*, a homolog of *LINC00152

As previously reported [40], *LINC00152* is a close homolog of another lncRNA on chromosome 2, *MIR4435-2HG*, both have nearly identical sequences (with only 6 base mismatches) and both contain M8 sequence (Fig 7A). *LINC00152* and *MIR4435-2HG* are both transcribed from chromosome 2, *LINC00152* is located at chr2: 87455476-87606739 and *MIR4435-2HG* is transcribed from chr2:111196350-111495115. In order to estimate the expression level of these two RNAs, we considered RNA-seq reads that uniquely mapped without any mismatch to either *LINC00152* or *MIR4435-2HG*. This analysis showed that *LINC00152* and *MIR4435-2HG* are expressed at the same level in U87 cells, and both of the RNAs are knocked down upon treatment of si*LINC00152* to a similar extent (Fig 7B and C). Thus, the phenotype that we observe with siRNA directed towards *LINC00152* is also likely through knocking down the highly similar *MIR4435-2HG*. Moreover, given the high similarity between the two transcripts, and the fact that, from nucleotides 382 to 478, *MIR4435-2HG* forms a 97 nucleotides long stem-loop in

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

the same position as *LINC00152*, it is likely that *MIR4435-2HG* overexpression phenocopies the effects of *LINC00152* on cell invasion. However, we see an increase in cell invasion when we exogenously express *LINC00152*, saying that *LINC00152* by itself can promote cell invasion. Again, upon considering uniquely mapped reads in TCGA RNA-seq data, we found that both *LINC00152* and *MIR4435-2HG* are equally expressed in LGG and GBM (Fig 7D). Analysis of TCGA RNA-seq data also revealed a positive correlation between the expression of the two RNAs in GBM and LGG (Fig 7E and G), suggesting that these two RNAs may be co-regulated. Moreover, the Kaplan Meier plot to estimate survival showed that expression of either RNA is associated with poor patient survival (Fig 7G and H).

Discussion

The human genome was once thought to be mainly dormant and that most of the transcription was devoted in 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 [41-43]. 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 [43-46]. 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 prostate cancer [11,14,47].

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

In this study we have shown that *LINC00152* is a lncRNA 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.

To assess the coding probability of *LINC00152* we used the Coding Potential Assessment Tool (CPAT) and found a score of 0.0289, suggesting that this lncRNA has no coding potential, since a CPAT score of < 0.5242 is considered non-coding [48]. In addition, ExPasy [49] predicts the first *LINC00152* ORF of 42 amino acids (126 nucleotides) from nucleotides 64 to 189. This ORF is not similar to any ORF predicted for the mouse transcript Gm14005. Moreover, blastx of the translated 126 nucleotide ORF did not show any hits in mouse protein database. In a different frame there is a longer ORF of 92 amino acids (which also does not have a homolog in the mouse transcript) that is preceded by a short ORF that has three stop codons. So, the longer ORF is also unlikely to be translated.

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Previous studies have shown that *LINC00152* is an oncogenic lncRNA involved in regulating invasion in different types of tumors [5,6,18-20] , including gliomas [5,6]. Mingjun Yu and collaborators [5] have reported an *in vivo* tumor xenograft study, downregulation of *LINC00152* produced smaller tumors and increased survival rates when compared to control. Thus, our findings reinforce the idea that *LINC00152* is an oncogenic lncRNA that is associated with aggressive tumors by promoting cell invasion. Moreover, through analysis of global RNA structure mapping and RNA-protein interaction data, we identified a protein bound stem-loop in the 3' region of *LINC00152*. The structure-function analysis demonstrated that this stem-loop is necessary and sufficient for stimulating invasion of U87 cells, that it can rescue the loss of invasion seen after knockdown of *LINC00152* and that the base-pairing of the opposite strands of the stem-loop, rather than the sequence at the mutated sites, is more important for stimulating invasion of U87 cells. However, it is likely that there are specific sequences along the M8 stem-loop that are important for *LINC00152* function that will be examined in a future study.

GSEA of RNA-seq from *LINC00152* knocked down cells also supports the idea that *LINC00152* is involved in promoting invasion. More specifically, *TPM2*, *PTX3*, *IGFBP4*, *TGM2*, *SPP1* and *LUM* were downregulated by si*LINC00152* and upregulated after *LINC00152* overexpression. Since we did not compare the global gene-expression changes with *LINC00152* overexpression and knockdown by RNA-seq, there are likely to be many other genes that will be regulated similarly to the six genes we tested in this study. Because si*LINC00152* decreased invasion, we focused on genes in the RNA-seq data whose change (up or down) will decrease invasion. Of these 13 transcripts qRT-

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

PCR after si*LINC00152* validated the changes in 12. Out of these 12, 6 genes were changed in the opposite direction when *LINC00152* full length or M8 was overexpressed (Sup Table 1).

In addition, despite previous suggestions, analysis of the RNA-seq showed us that *LINC00152* is not acting as a ceRNA that sponges miRNAs.

LINC00152 is expressed from a syntenic location from mouse transcript. Mouse has a single gene, *MIR4435-2HG* (*Gm14005* or *MORRBID*), but humans have two closely related genes, *LINC00152* and *MIR4435-2HG*.

MIR4435-2HG is a host gene for a miRNA, as miR-4435 is transcribed from an intron of *MIR4435-2HG*. However, none of the effects observed by overexpression of *LINC00152* are due to miR-4435, since the functional assays were done using the cDNA of *LINC00152*. Furthermore, miR-4435 is not detected in any of the short RNA-seq libraries (such as miRGator v3.0 and miRmine) from brain, glial cell lines and gliomas.

Mouse *MIR4435-2HG* has been proposed to be a pro-survival lncRNA that represses a gene *in cis*, the proapoptic gene *BCL2L11* (*BIM*) by recruiting the polycomb repressive complex, PRC2, to the *BCL2L11* promoter [50]. We considered the possibility that *LINC00152*, although cytoplasmic, is acting as a pro-oncogenic RNA by similarly suppressing BCL2L11. si*LINC00152* increases *BCL2L11* RNA (Sup Fig 6B-D), but this is not expected to decrease cell invasion. Second, overexpression of *LINC00152* from a heterologous site or the M8 hairpin of the lncRNA did not decrease *BCL2L11* (Sup Fig 6E-G) and yet increased cell invasion. Third, analyzing previously published mouse PAR-CLIP data, we determined that EZH2 from the polycomb complex associates with

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

an intronic region of *MORRBID* (mouse *MIR4435-2HG*/MORRBID) which is not near the M8 region. Fourth mouse *MIR4435-2HG* encodes only the first third of the M8 hairpin that we have found to be functions in human *LINC00152* or human *MIR4435-2HG*. Finally, *LINC00152* is predominantly cytoplasmic, arguing against any role in recruiting any factors to the genome. Collectively, these results suggest that interaction with PRC2 is not necessary for the stimulation of invasion seen upon overexpression of the *LINC00152* or the M8 hairpin RNA.

In conclusion, *LINC00152*/*CYTOR* and its homolog *MIR4435-2HG* functions as an oncogenic lncRNA in GBMs through the action of a protein-bound stem-loop and potentially plays a critical oncogenic role in a wide variety of cancer types. The results rule out a mechanism of action involving the sponging of miRNAs as proposed in the literature, or interaction with the Polycomb complex proposed for the mouse *MIR4435-2HG*/MORRBID RNA. *LINC00152* could also serve as a tumor biomarker or a target for future cancer therapeutics.

Acknowledgments

We thank Dutta lab members for advice and helpful discussions. This work was supported by grants from the NIH R01 CA166054, AR067712 and a V foundation award D2018-002 to AD. BTRK was partly supported by CAPES BEX 0320-13-7. MK was partly supported by a DOD award PC151085.

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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. 2015;17(suppl 4):iv1-iv62.

2. Huse JT, Holland EC. Targeting brain cancer: advances in the molecular pathology of malignant glioma and medulloblastoma. Nat Rev Cancer. 2010;10(5):319–31.

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

4. 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.

5. Yu M, Xue Y, Zheng J, Liu X, Yu H, Liu L, et al. Linc00152 promotes malignant progression of glioma stem cells by regulating miR-103a-3p/FEZF1/CDC25A pathway. Mol Cancer. 2017;16(1):110.

6. Zhu Z, Dai J, Liao Y, Ma J, Zhou W. Knockdown of Long Noncoding RNA LINC0000125 Suppresses Cellular Proliferation and Invasion in Glioma Cells by Regulating MiR-4775. Oncol Res. 2017. doi: 10.3727/096504017X15016337254597.

7. 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. 2009;10(5):459–66.

8. Consortium TEP. An Integrated Encyclopedia of DNA Elements in the Human Genome. Nature. 2012;489(7414):57–74.

9. 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.

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. 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. 2013;45(11):1392–8.

12. 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. 2013 Feb 1;73(3):1180–9.

13. 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.

14. 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. 2012;56(5):1817–27.

15. Cao WJ, Wu HL, He BS, Zhang YS, Zhang ZY. Analysis of long non-coding RNA expression profiles in gastric cancer. World J Gastroenterol. 2013;19(23):3658–64.

16. 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. 2016;7(52):85728-85740.

17. Zhang YH, Fu J, Zhang ZJ, Ge CC, Yi Y. LncRNA-LINC00152 down-regulated by miR-376c-3p restricts viability and promotes apoptosis of colorectal cancer cells. Am J Transl Res. 2016;8(12):5286-5297.

18. Teng W, Qiu C, He Z, Wang G, Xue Y, Hui X. Linc00152 suppresses apoptosis and promotes migration by sponging miR-4767 in vascular endothelial cells. Oncotarget. 2017;8(49):85014-85023.

19. Cai Q, Wang Z, Wang S, Weng M, Zhou D, Li C, et al. Long non-coding RNA LINC00152 promotes gallbladder cancer metastasis and epithelial-mesenchymal transition by regulating HIF-1α via miR-138. Open Biol. 2017;7(1). pii: 160247.

20. Wang Y, Liu J, Bai H, Dang Y, Lv P, Wu S. Long intergenic non-coding RNA 00152 promotes renal cell carcinoma progression by epigenetically suppressing P16 and negatively regulates miR-205. Am J Cancer Res. 2017;7(2):312-322.

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. 2016;13(12):e1002192.

22. 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. 2012;2(5):401 LP-404.

23. 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. 2013;6(269):pl1 LP-pl1.

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

25. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Meth. 2015;12(4):357–60

26. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2014;31(2):166–9.

27. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

28. 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. 2005;102(43):15545–50.

29. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31(13):3406-15.

30. 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. 2016;165(5):1267–79.

31. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.

32. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotech. 2011;29(1):24–6.

33. Ji Z, Song R, Huang H, Regev A, Struhl K. Transcriptome-scale RNase-footprinting of RNA-protein complexes. Nat Biotech. 2016;34(4):410–3.

34. Network CGAR. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. N Engl J Med. 2015;372(26):2481–98.

35. 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. 2015;6(40):42813–24.

36. 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. 2015;14(19):3112–23.

37. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science (80-). 2009;324(5924):218 LP-223.

38. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. Cell. 2013;154(1):240–51.

39. 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. 2014;34(33):10924–36.

40. Nötzold L, Frank L, Gandhi M, Polycarpou-Schwarz M, Groß M, Gunkel M, Beil N, Erfle H, Harder N, Rohr K, Trendel J, Krijgsveld J, Longerich T, Schirmacher P, Boutros M, Erhardt S, Diederichs S. The long non-coding RNA LINC00152 is essential for cell cycle progression through mitosis in HeLa cells. Sci Rep. 2017;7(1):2265.

41. Mirza AH, Kaur S, Brorsson CA, Pociot F. Effects of GWAS-Associated Genetic Variants on lncRNAs within IBD and T1D Candidate Loci. PLoS One. 2014;9(8):e105723.

42. Huarte M. The emerging role of lncRNAs in cancer. Nat Med. 2015;21(11):1253–61.

43. 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.

44. 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. 2013;24(2):206–14.

45. 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.

46. 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. 2016;2(9):e1600220.

47. 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. 2007;132(1):330–42.

48. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013 Apr 1;41(6):e74.

49. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003 Jul 1;31(13):3784-8.

50. Kotzin JJ, Spencer SP, McCright SJ, Kumar DBU, Collet MA, Mowel WK et al. The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. Nature. 2016;537(7619):239-243.

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

FIGURE LEGENDS

Figure 1. *LINC00152* is upregulated in aggressive gliomas and in many cancer types. A) Boxplot of *LINC00152* expression in normal brain tissue, G2 (grade 2 glioma), G3 (grade 3 glioma) and GBM. B) Boxplot of *LINC00152* expression in LGG subtypes and normal brain tissue. C – L) 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 2. High level of *LINC00152* expression is associated with poor patient prognosis in GBMs, LGGs and many other tumors. A) Kaplan Meier of GBM patients separated into the top 33% highest expressing *LINC00152* cohort and lower expressing cohort. B) Kaplan Meier of LGG patients separated into the 33% highest expressing *LINC00152* cohort and lower expressing cohort. C-F) 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. Hazard ratio is indicated as "HR" and the 95% confidence interval is indicated as "CI".

Figure 3. *LINC00152* is a cytoplasmic lncRNA that promotes cell invasion in U87 cells. A) Western blot of Lamin A/C and Actin, markers of the nucleus and cytoplasm, respectively. B) qRT-PCR of *LINC00152* and a cytoplasmic RNA marker, GAPDH, and a nuclear RNA marker, MALAT1. C) In-situ hybridization of *LINC00152* in U87 and U251 cell lines; DRAIC lncRNA probes were used as negative control ("- control").

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Purple color: positive signal. D) qRT-PCR showing knockdown of *LINC00152* after treatment with two different siRNAs. E) Invasion assay with U87 cells after treatment with two different siRNAs against *LINC00152*; * p-value < 0.05. Pictures were adjusted by -20% in brightness and +40% in contrast. F) qRT-PCR showing overexpression of *LINC00152* after transient overexpression. G) Invasion assay with U87 cells overexpressing *LINC00152*; * p-value < 0.05. Pictures were adjusted by +40% in contrast.

Figure 4. *LINC00152* regulates genes involved in invasion in U87 cells. A) Volcano plot of statistical significance against fold-change highlighting differentially regulated genes in black color upon si*LINC00152* in U87 cells. B) Plot from gene set enrichment analysis (GSEA) showing the gene set involved in epithelial-to-mesenchymal transition (EMT) enriched among upregulated genes (left black end of spectrum) after *LINC00152* knockdown in U87 cells. C) Cumulative distribution frequency plots of miRNA target mRNAs (as predicted by TargetScan: black line) or non-targets (grey line) showing fraction of genes with fold change less than that indicated on the X-axis after *LINC00152* knockdown. None of the miRNAs previously proposed to be sponged by *LINC00152* are released as evident from the fact that their targets are not repressed upon *LINC00152* knockdown.

Figure 5. A 120 nucleotide hairpin at the 3' end of *LINC00152* (M8) is sufficient for promoting cell invasion in U87 cells. A) Predicted secondary structure of *LINC00152* and the stem loop and protein bound regions identified by PARIS (RNA Duplex) and Ribo-seq (Sup Fig 4). B) Schematics of *LINC00152* deletion mutants. C) *LINC00152* qRT-PCR confirming overexpression levels of the different constructs. D) Invasion of

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

U87 cells after overexpressing the different *LINC00152* deletion mutants; * p-value < 0.05. Pictures were adjusted by -10% in brightness and +40% in contrast. E) Invasion of U87 cells decreases after treatment with si00152_II but is rescued by *LINC00152* m8 overexpression; * p-value < 0.05. Pictures were adjusted by +20% in brightness and +40% in contrast.

Figure 6. Point-mutation of nucleotides 333-336 or 349-352 of *LINC00152* shows the importance of M8 hairpin for stimulating invasion. A) Invasion of U87 cells after the different *LINC00152* deletion mutants are overexpressed. Mut A or mut B are incapable of inducing invasion in U87 cells. Combining the two mutants (mut AB) restores the hairpin and induces invasion to the same level as full length *LINC00152*; * p-value < 0.05. Pictures were adjusted by +20% in brightness and +40% in contrast. B) qRT-PCR confirming overexpression of the different *LINC00152* constructs. C) Predicted secondary structure of full length *LINC00152* with the black line marking the sequence at the tip of the hairpin and the light grey and dark grey lines marking the residues that are mutated in Mut A or B, respectively. D) Predicted secondary structures of *LINC00152* mutants A, B and AB. The black, light grey and dark grey lines mark the corresponding residues as in Fig. 6C.

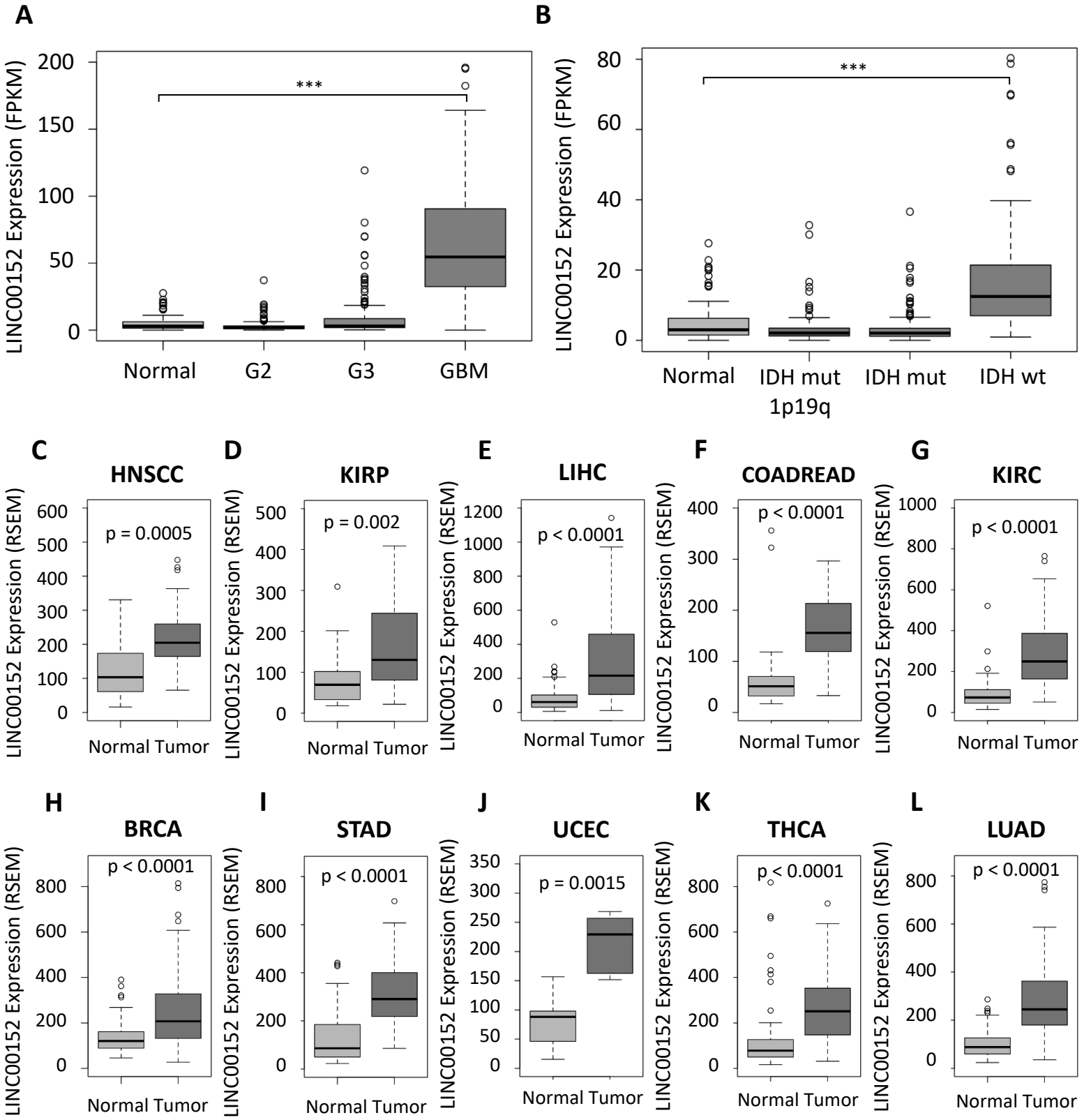
Figure 7. *LINC00152* is highly similar to the lncRNA MIR4435-2HG. A) Sequence alignment of *LINC00152* and MIR4435-2HG. M8 is highlighted in the boxed area. B) *LINC00152* specific RNA-seq reads in cells treated with siGL2 or si*LINC00152*. C) MIR4435-2HG specific RNA-seq reads in cells treated with siGL2 or si*LINC00152*. D) *LINC00152* or MIR4435-2HG specific reads in TCGA RNA-seq data for LGG and GBM. E) Correlation of expression of *LINC00152* and MIR4435-2HG in LGGs (spearman

Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

correlation 0.68 p value < 2.2 e-16). F) Correlation of expression of *LINC00152* and MIR4435-2HG in GBMs (spearman correlation: 0.86, P value < 2.2e-16). G) Kaplan Meier Plot of LGG patients separated into the 50% highest expressing *LINC00152* cohort and the lowest 50% expressing cohort. H) Kaplan Meier Plot of LGG patients separated into the 50% highest expressing MIR4435-2HG cohort and the lowest 50% expressing cohort. Hazard ratio is indicated as “HR” and the 95% confidence interval is indicated as “CI”.

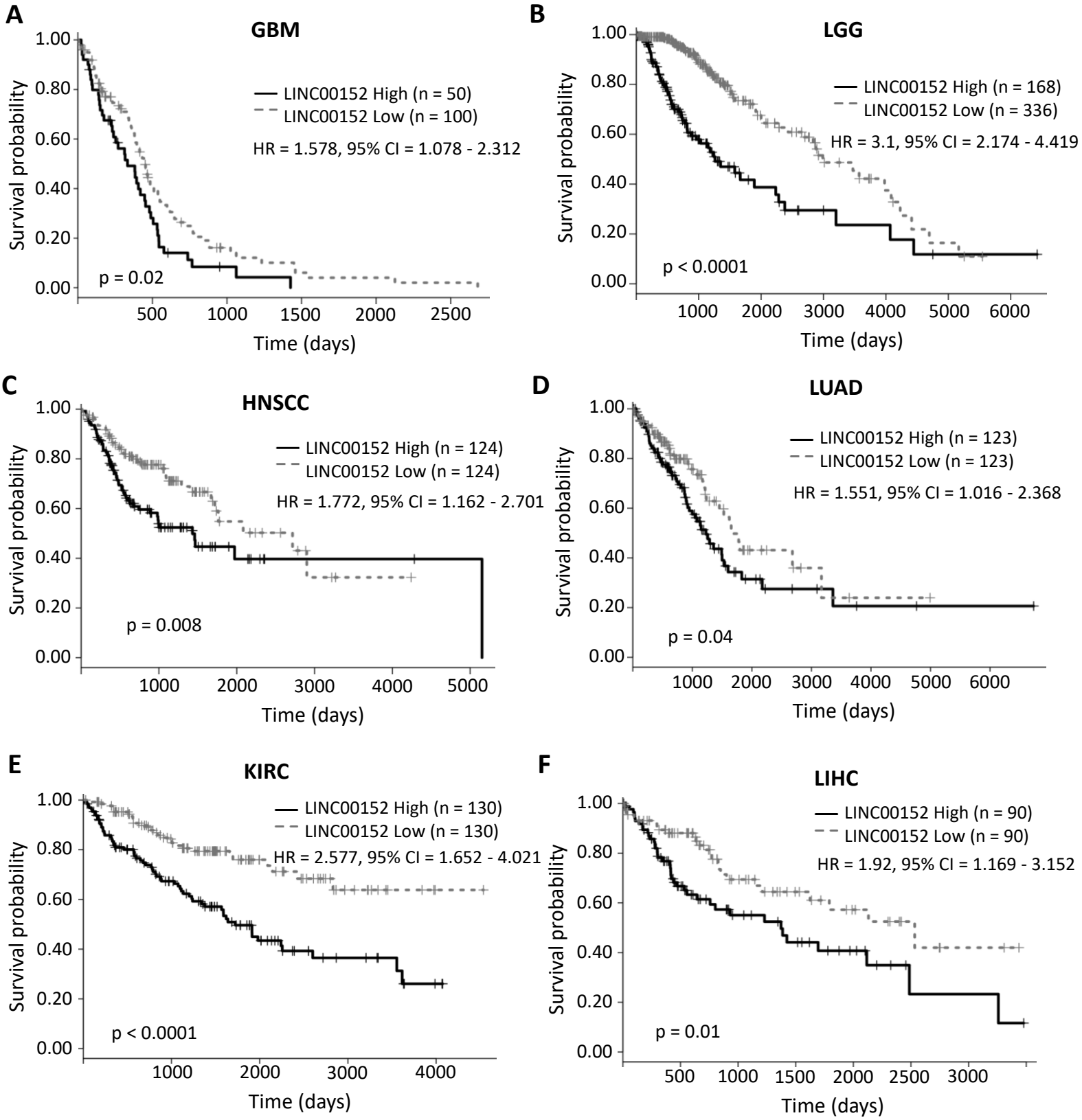
Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 1



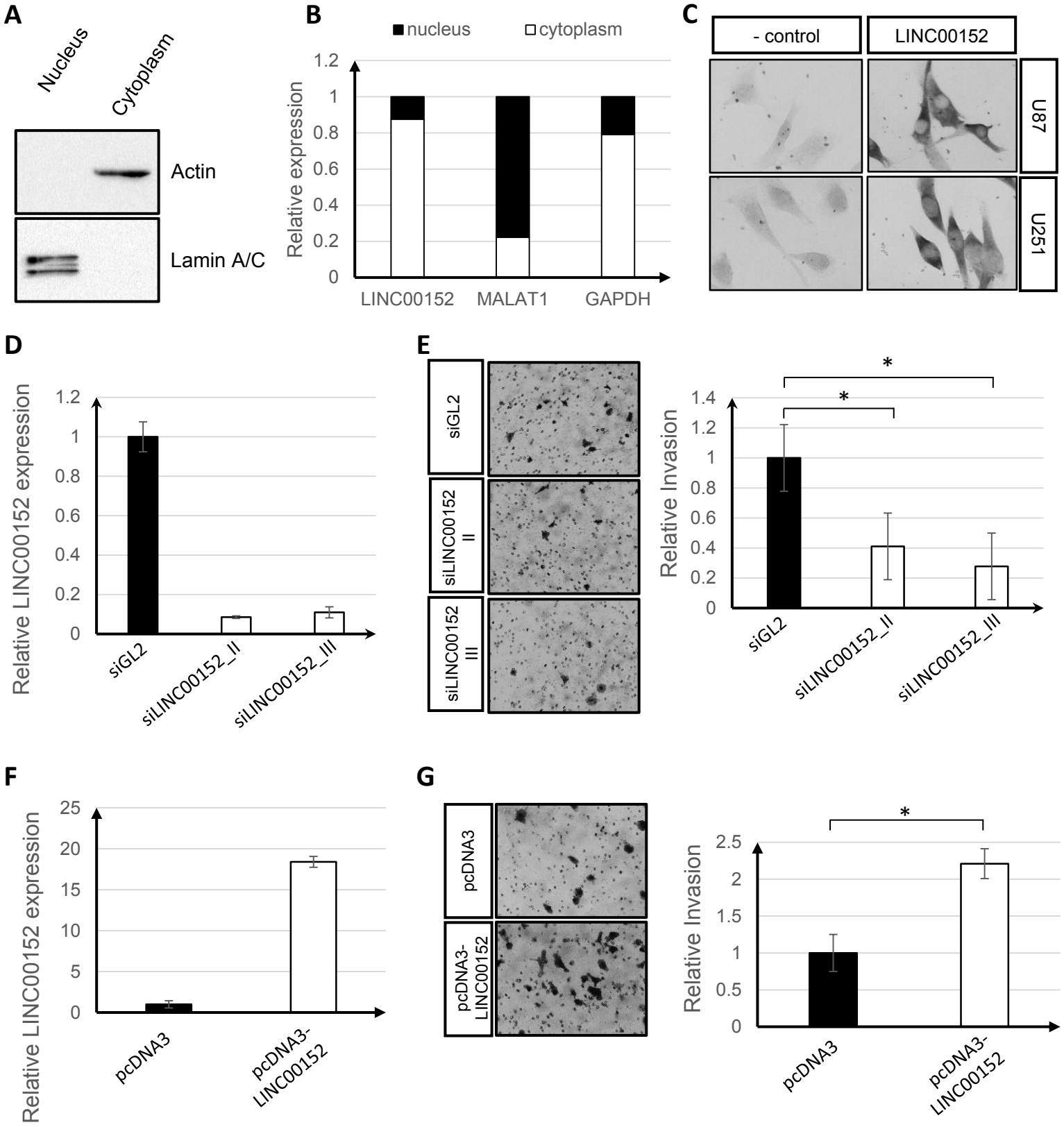
Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 2



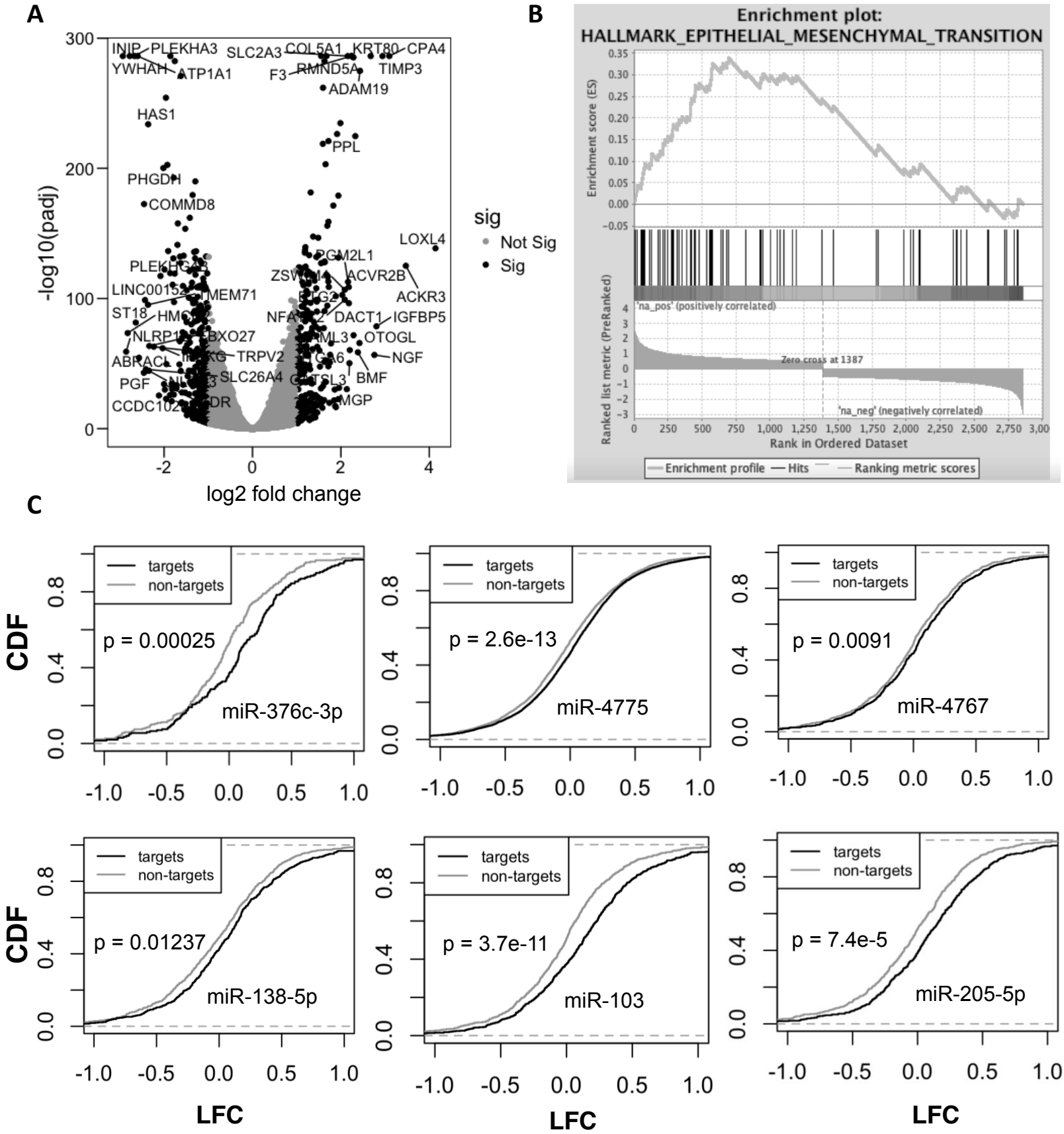
Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 3



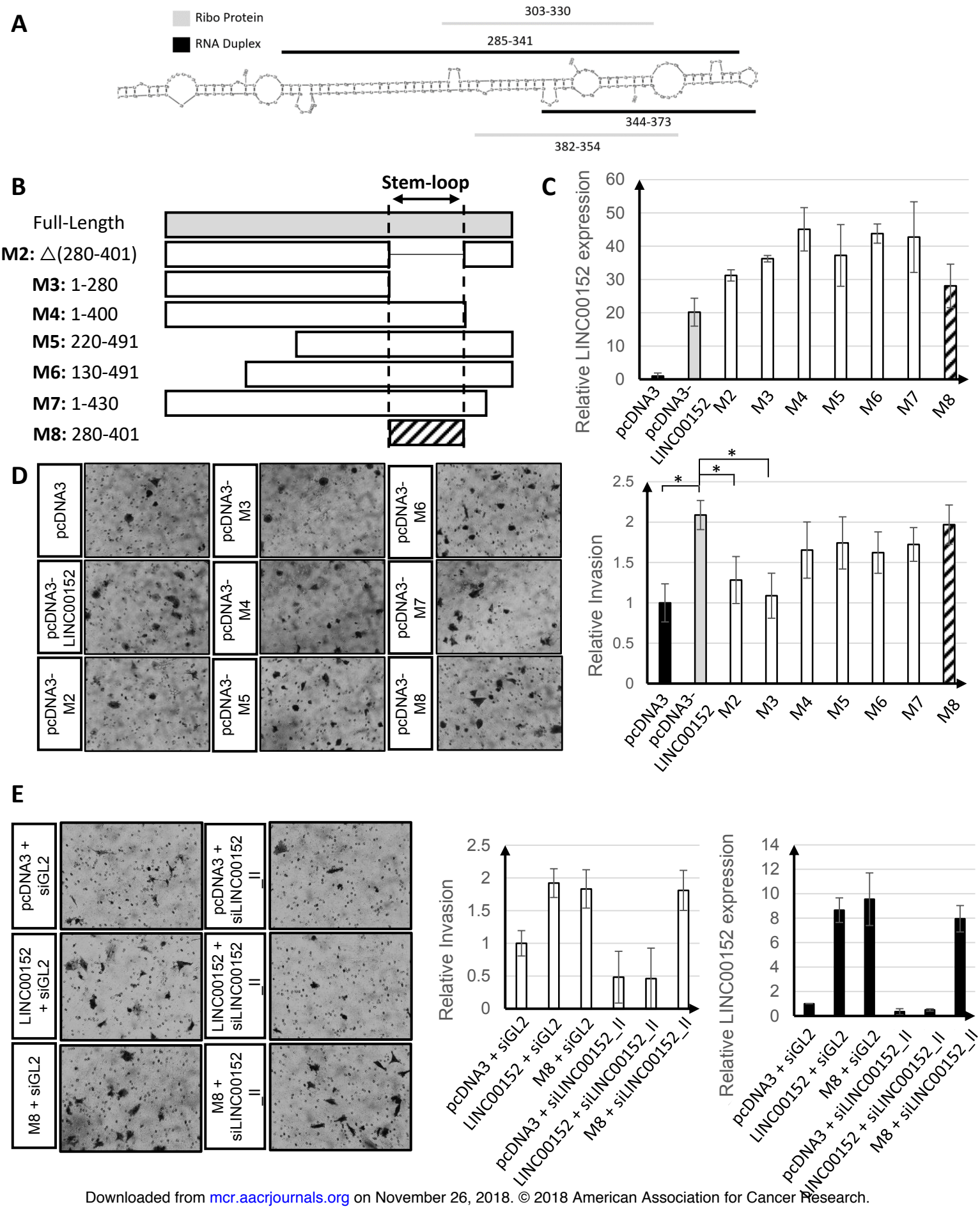
Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 4



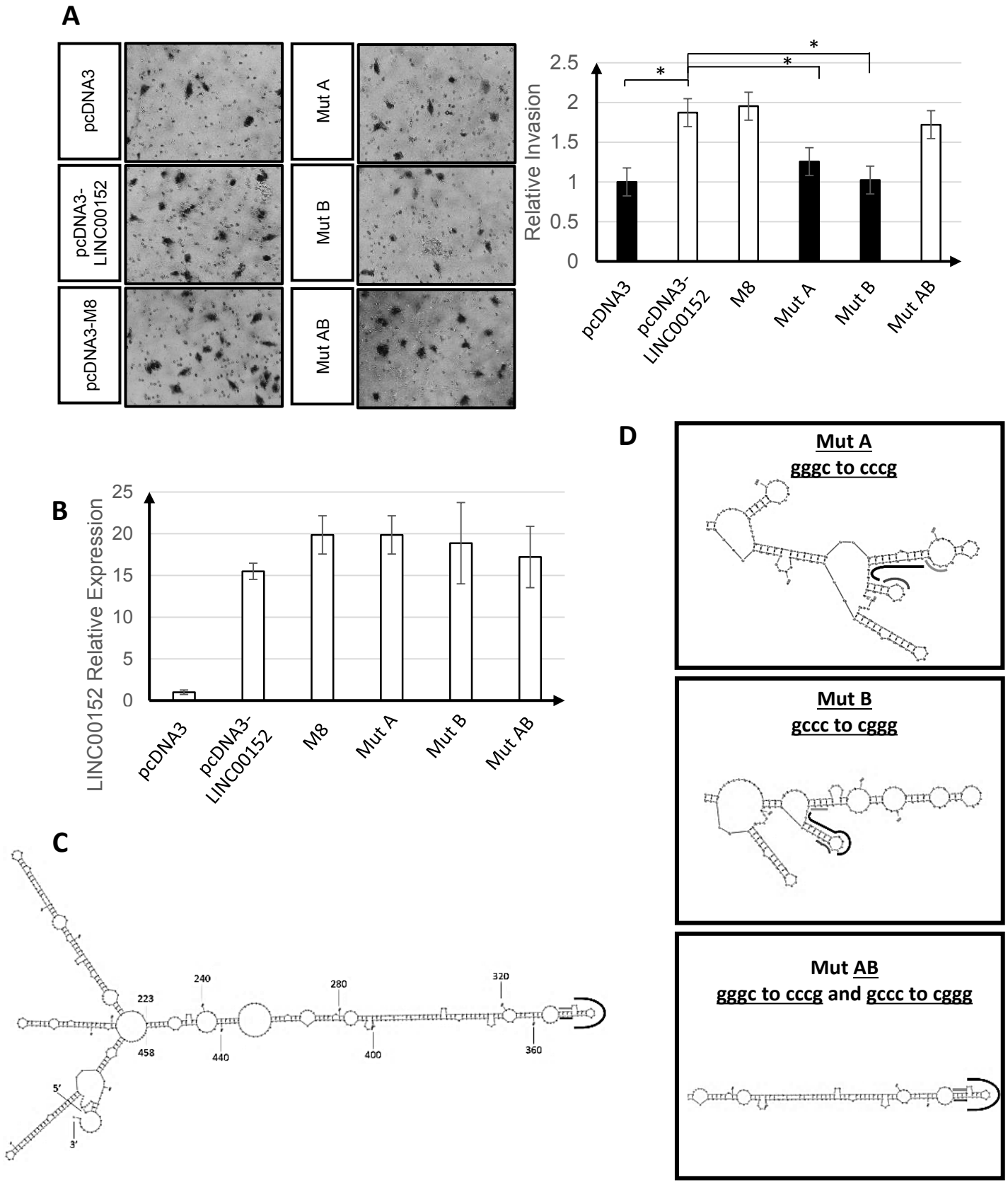
Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 5



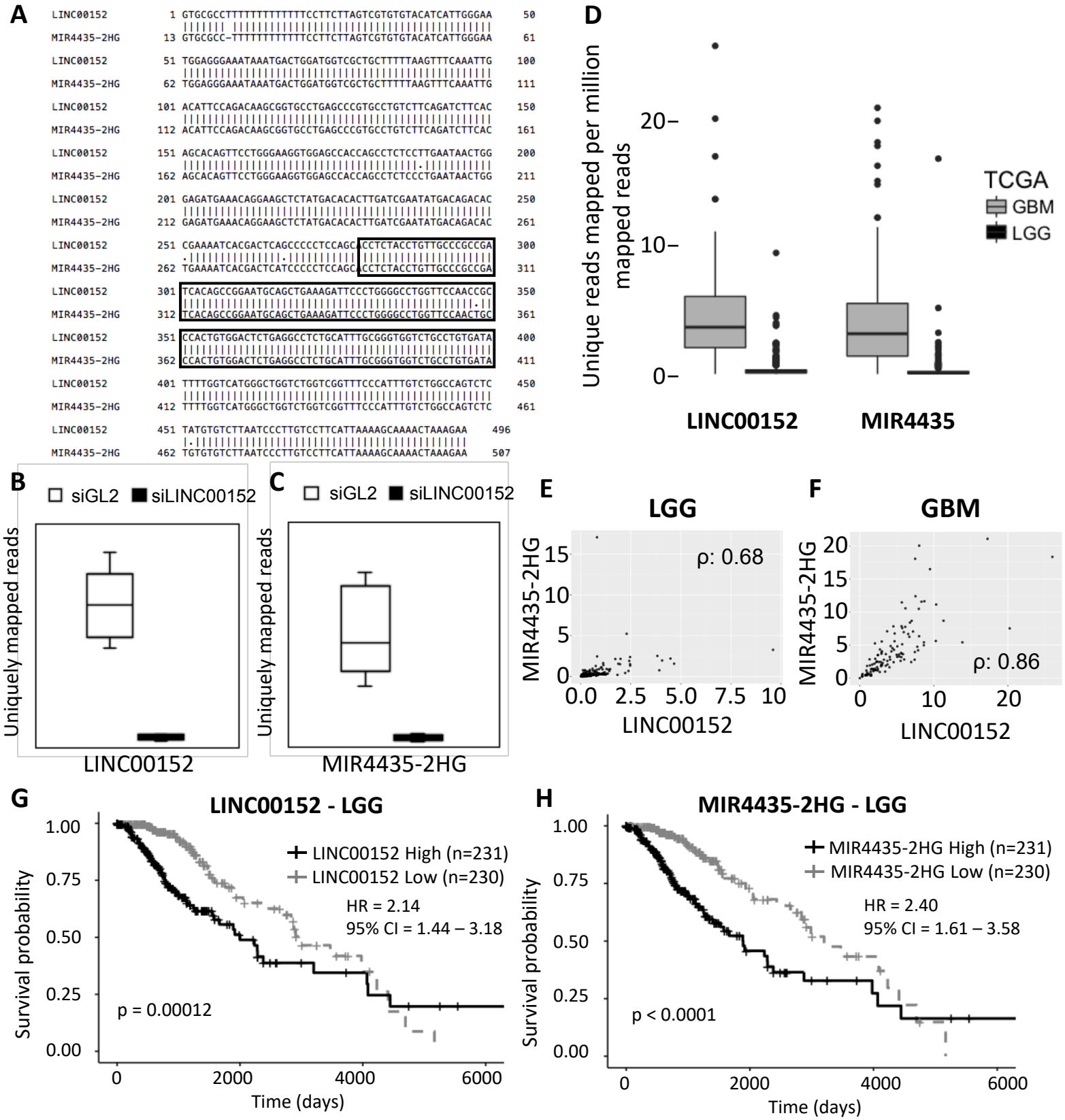
Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 6



Author Manuscript Published OnlineFirst on July 10, 2018; DOI: 10.1158/1541-7786.MCR-18-0322
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 7



AACR American Association for Cancer Research

Molecular Cancer Research

LINC00152 Promotes Invasion Through a 3'-hairpin Structure and Associates with Prognosis in Glioblastoma

Brian J. Reon, Bruno Takao Real Karia, Manjari Kiran, et al.

Mol Cancer Res Published OnlineFirst July 10, 2018.

Transfer RNA Fragments (tRFs): a Novel Class of Non-micro Short RNAs that Uses Ago1, 3 and 4 to Repress Specific Target RNAs Through 5′ Seed Sequences

Anindya Dutta, Pankaj Kumar, Manjari Kiran and Caman Kuscu

University of Virginia, Charlottesville, VA

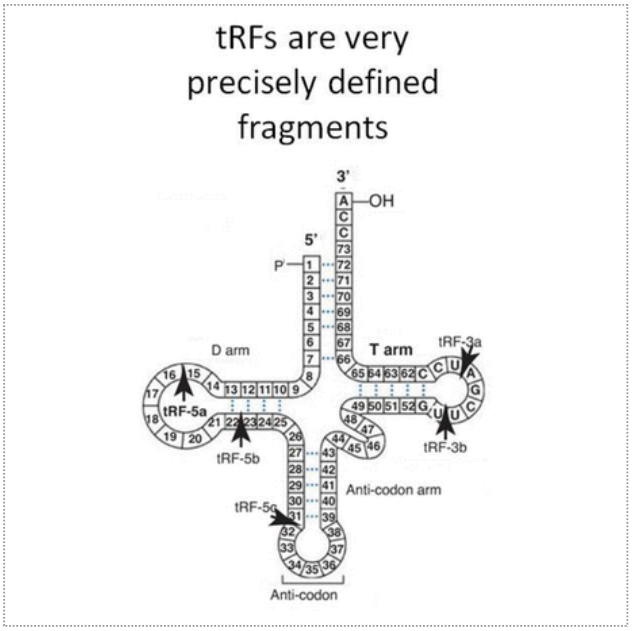
Abstract

tRFs, 14-32 nt long single-stranded RNA derived from mature or precursor tRNAs, are a recently discovered class of small RNA present at read counts comparable to miRNAs. The tRFs are precisely generated fragments present in all human cell lines, mice, flies, worms, yeasts and even some bacteria and originate from the 5′ end (tRF-5) or 3′ end (tRF-3) of mature tRNAs or from 3′ trailer sequences of primary tRNA transcripts (tRF-1). Genes involved in generating canonical miRNAs or siRNAs (Dicer or DGCR8) are dispensable for tRF generation. tRF-1s and tRF-3s are more abundant in the cytoplasm than the nucleus, but tRF-5s are enriched in the nucleus.

Human Ago PAR-CLIP data show that tRF-5s and tRF-3s associate with Ago-1, -3, and -4 rather than Ago-2 (unlike microRNAs), raising intriguing questions about how these single-stranded RNA fragments are loaded on to Ago complexes and how the selectivity is determined for Ago-1, -3 and -4 versus Ago-2. tRF-1s are not associated with Ago proteins. The locations of the U to C mutation caused by the cross-linking of the thio-uracil in the tRF or the target RNA to the Ago protein demonstrate that a 5′ seed sequence of 6-7 bases in tRF-5 or -3 is used to recognize the target RNA in exactly the same way as used by a microRNA to recognize its target RNA. Human Ago-1 CLASH data identify tens of thousands of chimeric tRF-target molecules produced by ligation of specific tRFs to a paired target RNA in the Ago-1 protein isolated from 293T cells. Surprisingly, tRF-target chimeras are 2-3 fold more abundant than microRNA-target chimeras suggesting that more tRFs than microRNAs are paired with targets in the Ago-1 complex. The chimeras identify hundreds of tRF targets and demonstrate that tRF-5s and tRF-3s in the Ago-1 complex use complementarity to their 5′ seed sequences to recognize the target RNAs. Expression of specific tRNAs predicted to produce the same tRF shows that not all tRNAs are equally proficient in producing a given tRF. tRF-3s produced from transfected tRNAs load into Ago complexes and specifically repress reporter genes with complementarity to the tRFs in their 3′ UTRs. Mutations of the target sites in the reporter or in the generating tRNA show that complementarity to the seed sequence of the tRF is critical for repression. In conclusion, tRF-5s and -3s are non-micro-short RNAs produced from specific tRNAs by Dicer-independent pathways to regulate gene expression.

Support or Funding Information

NIH P01 CA104106



Footnotes

This abstract is from the Experimental Biology 2016 Meeting. There is no full text article associated with this abstract published in The FASEB Journal.