UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Investigating the miRNA pathways contribution to intra-tumour heterogeneity in glioblastoma and RNA binding of isoforms of the miRNA effector protein Argonaute

by Christopher Michael Smith

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Gyorgy Hutvagner, Daniel Catchpoole, and Jinyan Li

University of Technology Sydney
Faculty of Engineering and IT

December 2022

Certificate of Original Authorship

I, Christopher Smith, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Biomedical Engineering, Faculty of Engineering and IT, at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Date: 2/12/2022

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my supervisors Gyorgy Hutvagner, Daniel Catchpoole, and Jinyan Li for their mentoring and continuous support throughout this degree. This PhD would not have been possible without you all.

A special thanks to Sarah Bajan for all her wet lab work, which generated much of the data used in this thesis, as well as her mentoring throughout my early years as a student. Thanks to Hanjie Wu, Chaowang Lan, and all other members of the Hutvagner lab for their suggestions and help in the lab and with bioinformatics. In addition, thanks to Diego Fajardon for setting me in the right direction during my first year when I was learning bioinformatics.

Over at the Children's Hospital at Westmead I would like to thank the members of the Tumour Bank, including Aysen Yuksel and Li Zhou, who helped me find my way around the lab and make sure I had everything I needed there. Lawrence Cantrell for his assistance with training so I could work with equipment in The Westmead Institute for Medical Research.

I would also like to thank Dan Carter and Dominik Beck for their insightful feedback in each of my candidature assessments.

Last but certainly not least, a special thanks to my family, including my partner Vee who was by my side through every moment during these years, and my parents for raising and supporting me so I could reach this point in my life.

Impact of the COVID-19 Pandemic on this Thesis

Our original plan to investigate our first hypothesis included an experiment to generate single cell RNA and small RNA data in a heterogeneous cancer model such as glioblastoma or neuroblastoma. This data was to be used to address both aims by enabling a direct comparison between miRNA and RNA expression profiles in potential cancer subpopulations. However, several months into the COVID-19 pandemic we decided to exclude this experiment, which we deemed high risk due to the scale and ambition of the project as well as the challenges in collaborating cross-institution during this period. We had concerns that access to researchers and laboratories would be impaired and make completion of this project unfeasible. In response, we retained our original aims but had to rely exclusively on public data. We chose glioblastoma as a heterogenous cancer model as it is one of the most extensively sequenced cancers with expression data available for RNAs and small RNAs in bulk tumours and single cells.

We also planned and initiated an experiment in 2021, approximately one year prior to my end of candidature. This involved collaboration with the Tumour Bank, part of The Children's Cancer Research Unit (CCRU) at The Children's Hospital at Westmead. This experiment intended to biologically validate observations of heterogeneity with miRNAs from the Dlk1-Dio3 locus in glioblastoma, by using a fluorescent in-situ hybridization assay (miRNAscope) for detection of miRNAs in formalin-fixed paraffin embedded glioblastoma samples. The project at the Tumour Bank was initially planned to begin late March in 2021. However, due to supplier errors and delays in shipping we were not able to receive all the required reagents to proceed with the experiment until June 2021, right before the second lockdown was instated in Sydney. The proximity of the outbreaks to Westmead hospital and the challenges in coordinating work with other lab members for assistance, access to facilities, and equipment installation, meant I was forced to delay this project until September 2021 with only a few months remaining before my final candidature assessment. While our preliminary results indicated that we were able to detect RNA using the control probes, the staining was inconsistent across cells, and we were unable to resolve this issue through optimization in the available timeframe.

We also had planned a second experiment in 2021, at the University of Technology, Sydney, to sequence Argonaute 2 using Nanopore sequencing for full length isoforms. This was intended to support the findings in chapter 4 where we used Capture-seq and short read sequencing to detect multiple isoforms of Argonaute 2 in human tissues and cell lines. To make this project feasible, we developed a modified library preparation protocol which could generate Argonaute 2 cDNA through targeted amplification. However, we were met with significant challenges with generating full length

amplicons of this gene, and because of constant delays throughout 2021 our timeline for both experiments clashed, and we chose to prioritise the miRNAscope project at the Tumour Bank.

Despite the challenges imposed by COVID-19 I was able to publish 3 first author papers – 1 review and 2 research papers.

Publications

Smith, C. M., Catchpoole, D. & Hutvagner, G. Non-Coding RNAs in Pediatric Solid Tumors. *Front. Genet.* **10**, 1–18 (2019).

Smith, C. *et al.* Cataloguing the small RNA content of honey using next generation sequencing. *Food Chem. Mol. Sci.* **2**, 100014 (2021).

Smith, C. M. & Hutvagner, G. A comparative analysis of single cell small RNA sequencing data reveals heterogeneous isomiR expression and regulation. *Sci. Rep.* **12**, 2834 (2022).

Table of Contents

List of Figures and Tables

1. Literature Review

2. The role of miRNAs in glioblastoma tumour heterogeneity and cell state regulation

Tables **Page**

3. A Comparative Analysis of Single Cell Small RNA Sequencing Data Reveals Heterogenous IsomiR Expression and Regulation

Figures **Page**

Tables **Page**

4. Investigating the RNA binding properties of alternatively spliced Argonaute 2 (Ago2) isoforms

Figures **Page**

Tables **Page**

Appendix

Figures **Page**

Tables **Page**

Abstract

miRNAs are highly abundant small non-coding RNAs that are essential for post-transcriptional gene regulation. Many of the mechanisms which regulate miRNA expression and function are poorly understood. Their dysregulation is documented extensively in many diseases including cancer.

Glioblastoma is a highly aggressive and heterogeneous brain cancer that affects patients of all ages. Intra-tumoural heterogeneity describes the existence of genomically distinct subpopulations of tumour cells which can lead to differences in growth rate, metastatic potential, or vulnerability to certain treatments. The first part of my thesis investigates how miRNAs and miRNA variants (isomiRs) may be involved in intra-tumoural heterogeneity by applying bioinformatics analyses to single cell small RNA and RNA sequencing data generated from previous studies. This work identified two miRNA clusters, the Dlk1-Dio3 locus and miR-224/452, as potential contributors to intra-tumour heterogeneity in glioblastoma and may be involved in cell state regulation. Additionally, we found evidence of cell autonomous regulation and function of isomiRs, highlighting another regulatory mechanism that may play a role in heterogenous cancers. These miRNAs may have utility as cancer biomarkers and implicate a novel set of targets for therapeutic research.

The second part of my thesis investigates splice variants (isoforms) of Argonaute, an essential protein in the miRNA pathway that mediates their regulatory effects. Numerous Argonaute isoforms have been previously annotated, with alterations in protein domains critical for miRNA binding and function. However, current studies base their assumptions of miRNA activity through a single variant of Argonaute and the consequence of these alterations and their biological relevance has not been investigated yet. We identified two variants of Argonaute with altered miRNA binding characteristics that are variably expressed in normal and cancerous cells, revealing a novel form of miRNA regulation that could also have implications in cancer research.

1. Literature review

1.1. miRNA biogenesis

1.1.1. miRNA biogenesis via the canonical pathway

For most miRNAs, their life cycle begins with transcription from DNA into a primary miRNA (pri-miRNA) by RNA polymerase II (Figure 1.1)[1]. Pri-miRNAs share several similarities to messenger RNAs (mRNAs) in that they are 5´ capped, 3´ polyadenylated, and can be several hundreds or thousands of nucleotides long. In many cases, the pri-miRNA encodes for one miRNA species, however in humans a substantial number are polycistronic and encode several different miRNAs. Pri-miRNAs must first be processed in the nucleus by the Microprocessor, a trimeric complex formed by the RNAse III enzyme Drosha and two DiGeorge Syndrome Critical Region 8 (DGCR8) molecules[2]. Pri-miRNAs contain imperfect stem-loop secondary structures which are recognized by DGCR8's double stranded binding domains. DGCR8 facilitates the anchoring of Drosha to the stem-loop structure, and Drosha then acts as a molecular ruler which cleaves the pri-miRNA around 11-13 nucleotides above the base of the stem. This releases a shorter 65-70 nt precursor RNA (pre-miRNA) with a 2 nucleotide overhang on the 3' end[3–5]. Sequence motifs at the base (UG and CNNC) and loop (UGU) structures are recognized by Drosha and DGRC8 respectively, aiding in the correct orientation and processing by the Microprocessor complex[2,6].

Translocation of the pre-miRNA from the nucleus to the cytoplasm is mediated by the transporter protein Exportin-5 (XPO5). XPO5 recognition and subsequent export of the double-stranded pre-miRNA is dependent on Ran-GTP[7].

In the cytoplasm the pre-miRNA is further processed by Dicer, another enzyme harbouring RNAse III domains, which specifically recognises both the pre-miRNA's loop structure and 2 nucleotide overhang[8]. Dicer acts as a second molecular ruler, cleaving approximately 21-25 nucleotides upstream from the base of the stem, which removes the loop structure to produce a double stranded miRNA duplex. This duplex contains two RNA strands named the 5 prime (5p) and 3 prime (3p) miRNA strands, based on which end of the precursor they originate from. Dicer is known to associate with other cofactors which can influence pre-miRNA processing. This includes proteins such as the double stranded RNA binding proteins TRBP and PACT, which have been shown to increase the precision of Dicer cleaving and alter the sequence of some of its products, as well Argonaute 2 (Ago2) which participates in the processing of certain pre-miRNAs.[9–11].

After Dicer processing, the double stranded miRNA duplex is incorporated into an Argonaute protein, which retains one of the strands to become the biologically active mature miRNA (guide strand) and ejects the other strand (passenger strand) for subsequent degradation[12]. This double stranded

complex is called the pre-RNA induced silencing complex (pre-RISC). Strand selection is determined by base pairing at the ends of the miRNA duplex, which influence which side of the duplex is less thermodynamically stable. This destabilization leads to the preferential loading of the strand with their 5' end exposed at this side.

The process of duplex loading and strand displacement requires coordination between several proteins and involves multiple functional domains on Argonaute[13]. During duplex loading, the ATP-dependent chaperone proteins Hsc70/Hsp90 are recruited to open up binding pockets in Argonaute and enable the large miRNA duplex to be anchored into the protein[14]. The duplex is then 'wedged' open at the 3' end of the guide strand, leading to ejection of the passenger strand from the complex[15].

The final product is a complex with the mature miRNA and Argonaute, commonly referred to as the functional or mature RNA-induced silencing complex (mature RISC), which is the minimal component for post-transcriptional miRNA-mediated gene regulation[13].

1.1.2. Non-canonical pathways for miRNA biogenesis

Besides the canonical miRNA pathway, several other pathways of miRNA biogenesis exist which do not involve Drosha and/or Dicer. There are several Drosha-independent pathways that could produce miRNAs and miRNA-like RNAs such as mirtrons, snoRNA and tRNA-derived miRNAs[16,17]. Mirtrons are miRNAs originating from the introns of messenger or non-coding RNAs, produced during splicing[16]. Although spliced introns are typically released as lariats, mirtrons can be linearised by the lariat debranching enzyme, enabling them to adopt a pre-miRNA structure for recognition by XPO5 and subsequent processing by Dicer in the cytoplasm. Mirtrons can either span the entire length of the intron or be associated with the 5´ or 3´ ends, the latter recruiting additional factors to trim the mirtron to the appropriate length[16]. snoRNAs such as the small Cajal body-specific RNA ACA45 and the atypical box C/D snoRNA U3 have been shown to be processed into small RNAs that associate with Argonaute following Dicer processing, with some evidence supporting their participation in miRNA-mediated regulation[18,19]. Likewise, the tRNA-derived small RNA CU1276 was also reported to be processed by Dicer into a functional miRNA that is loaded into Argonaute[17].

The miRNA miR-451 is unique in that it is the only known miRNA to be Dicer-independent, as its precursor can be directly processed by Ago2 through its catalytically active domain. This property enables miR-451 to retain a high level of expression during erythropoiesis, as the canonical miRNA pathway is dampened via Dicer suppression[20,21].

1.1.3. Alternative splicing regulates the function of miRNA biogenesis proteins

Alternative splicing enables functional diversification of protein-coding genes by altering the coding sequences or untranslated regions of messenger RNAs[22,23]. Studies have shown that miRNA biogenesis proteins frequently undergo alternative splicing, in some cases impacting small RNA processing and leading to changes in post-transcriptional gene regulation[24–27].

For example, alternative splicing can affect the subcellular localization of Drosha by changing the coding sequence of the N-terminal domain, with some isoforms localizing to the cytoplasm[24,28]. A later study demonstrated that pri-miR-34a is transported to the cytoplasm of quiescent human foreskin fibroblasts by Exportin-1 and are subsequently processed by cytoplasmic-localized Drosha isoforms[29]. This highlighted the existence of an alternative miRNA biogenesis pathway which remains active during quiescence while the canonical pathway is suppressed.

In humans, a large number of alternatively spliced isoforms of Dicer have been identified with some variants expressed in specific tissues[26,30,31]. For example, two variants of Dicer were identified in multiple human breast cancer cell lines, with one variant harbouring a reduced propensity for translation due to the skipping of two exons[32]. Later studies, also with breast cancer cell lines, found differential expression of dicer variants in cells with a mesenchymal phenotype compared to those with an epithelial phenotype, suggesting the alternative splicing of Dicer may function within the epithelial-mesenchymal transition (EMT)[33]. Dicer isoforms may also have functional significance in cancer. A truncated Dicer lacking a double stranded binding domain was found to be expressed in neuroblastoma cells but was not identified in any normal tissues[26]. Additionally, another truncated Dicer isoform was highly expressed in oral cancer cells and was downregulated during EMT[31]. Furthermore, in mice, functional diversification of Dicer isoforms was demonstrated in oocytes where the miRNA pathway is normally suppressed. An isoform of Dicer which lacks its N-terminal DExD helicase domain, was dominantly expressed[27]. This isoform, in contrast to its full-length form, was far more efficient in processing long double stranded RNAs into small interfering RNAs.

Finally, a longer variant of Ago1, named Ago1x, was discovered in multiple human breast cancer cell lines which could bind miRNAs in a similar manner to the canonical Ago1, but could not participate in miRNA-mediated gene regulation due to a loss of interaction with GW182[34,35].

1.1.4. miRNAs often form clusters

In humans, approximately one quarter of annotated miRNAs (481 of 1917 in miRBase) exist in clusters containing multiple miRNA genes[36]. The majority of miRNA clusters contain less than 5 miRNAs; however the human genome contains much larger clusters such as the 14q32 and C19MC miRNA clusters which host approximately 53 and 46 pre-miRNA genes respectively[37,38]. Clustered miRNAs are

co-transcribed and are typically under the same regulatory control through shared promoters and epigenetic regulatory factors. miRNA clusters are highly conserved across species and have established roles in cell and tissue development, immune response, and regulation of the nervous system[39,40]. It is believed that miRNA clusters formed from genetic events including gene duplications and base mutations, and evolved to facilitate the coordination of critical biological pathways, as miRNAs acting in tandem can target an extremely large but selective pool of genes[41].

One of the most well studied miRNA clusters is the miR-17/92 cluster (previously named C13orf25), located on chromosome 13 within the MIR17HG gene. This cluster contains 6 different miRNAs – miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1 and miR-92a-1, which are initially co-expressed as a single transcript approximately 800 nt long before being processing into individual pre-miRNAs. The miR-17/92 cluster is regulated by many transcription factors which bind to the MIR17HG, such as MYC and the E2F family, which can induce miR-17/92 expression, as well as p53 which blocks transcription. Many of the transcription factors that are responsible for regulating the miR-17/92 cluster are in turn regulated by these same miRNAs, forming autoregulatory feedback loops which help prevent uncontrolled expression levels[42].

Ostensibly, the largest miRNA cluster is found in the chromosomal region of 14q32 within a region often referred to as the Dlk1-Dio3 locus[37]. With the exception of one, all miRNAs from this region are believed to be transcribed as part of a large polycistronic transcript approximately 200kb long[43]. Unlike the miR-17/92 cluster, this polycistronic transcript also contains non-coding RNAs other than miRNAs, including at least 3 long non-coding RNAs, MEG3, MEG8 and MEG9, as well as the small nucleolar RNA SNORD112, 9 paralogous copies of SNORD113 and 31 copies of SNORD114[44].

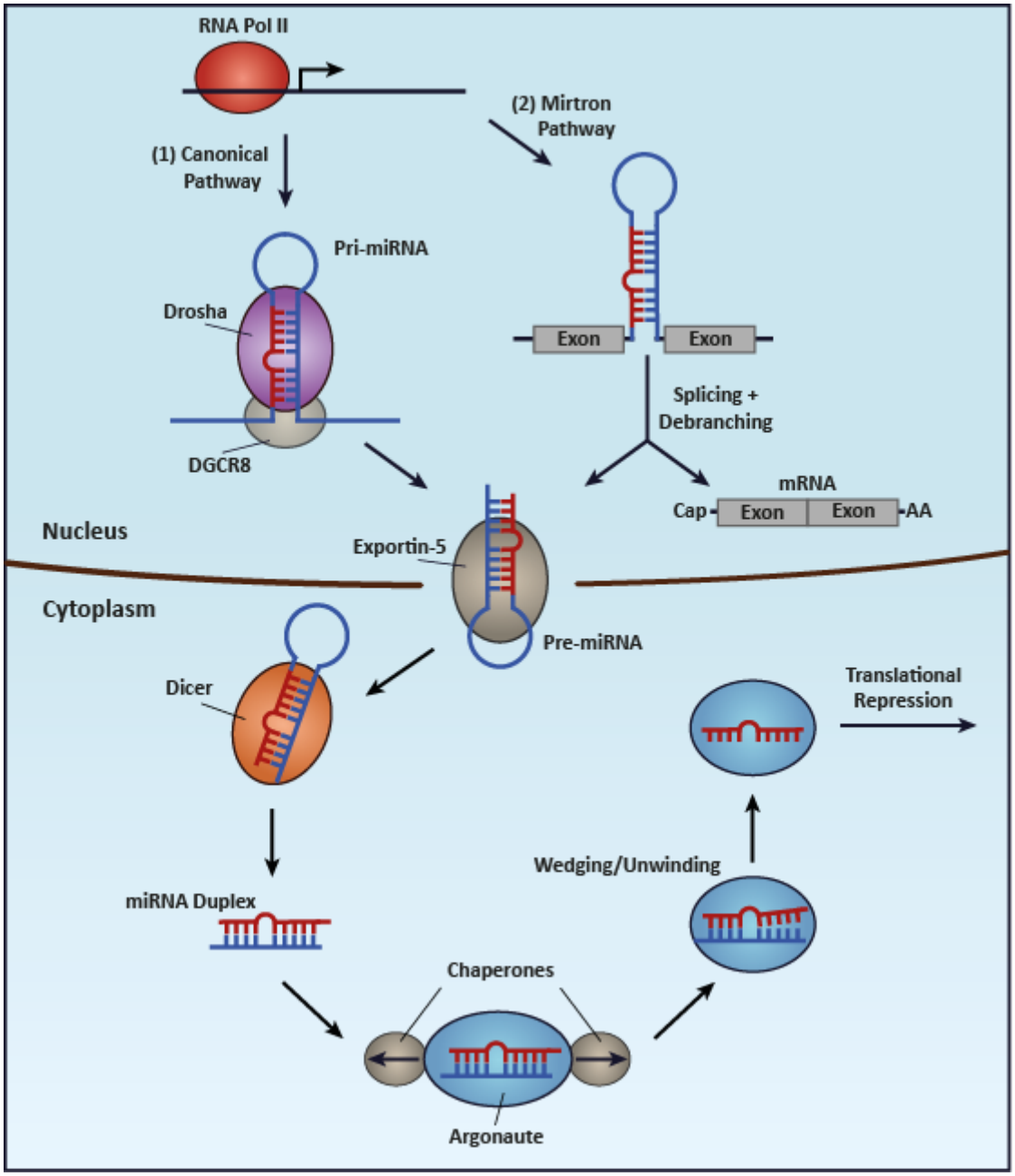

Figure 1.1. Pathways of miRNA biogenesis. miRNAs can originate from different biogenesis pathways. Two pathways, the canonical pathway and mirtron pathway are shown above. Canonical Pathway (1) – miRNAs are first transcribed as pri-miRNAs by RNA Pol II, which contains stem-loop structures recognizable by Drosha/DGCR8. Drosha/DGCR8 cleaves and liberates the stem-loop structure, which is then exported out of the nucleus and further processed by Dicer to produce a miRNA duplex. This duplex is then loaded into an Argonaute, which removes one strand and retains the other for translational repression. Mirtron Pathway (2) – precursor miRNAs are spliced out from the introns of mRNAs, which then re-join the canonical pathway after export into the cytoplasm. **Pri-miRNAs**: Primary microRNAs. **DGCR8**: DiGeorge Syndrome Critical Region 8. Figure from Smith et al, Honours, unpublished.

1.2. IsomiRs: a hidden layer of gene regulation

In eukaryotes, genes can be alternatively spliced to produce many isoforms with distinct biological functions or regulatory features[45]. In a similar manner, most miRNAs are expressed as different isoforms (referred to as isomiRs)[46]. IsomiRs can vary from their canonical sequences in at least four different ways (Figure 1.2). Firstly, 5´ or 3´ trimming variations can occur from imprecise processing by Drosha and Dicer. Variations caused by alternative trimming sites produce templated isomiRs, as the resulting sequence still matches the genomic sequence of the miRNA gene. IsomiRs can also result from untemplated modifications which are produced either by nucleotide additions at the 3´ end (tailing) of the mature or precursor miRNA or by nucleotide substitutions[47]. While most miRNA studies do not investigate isomiRs, an increasing number of studies have shown that these minor changes in sequence can lead to important changes in processing, function, or stability, and constitute an additional layer of gene regulation in the miRNA pathway[48].

1.2.1. Uridylation regulates the processing of several pre-miRNAs

Most pri-miRNAs are processed by Drosha into pre-miRNAs with a 2 nucleotide overhang on the 3´ end which enables efficient processing by Dicer[49,50]. However, several pre-miRNAs from the let-7 family of miRNAs and miR-105 are processed with 1 nucleotide 3´ overhangs and require extension via monouridylation in order to be processed efficiently into a miRNA duplex[51,52]. As the site of monouridylation is the 3´ end of the duplexes 3p strand, any 3p miRNAs which are not degraded will possess a 3´ non-templated uridine. Conversely polyuridylation can block pre-miRNA processing, which has been shown in cells expressing the Lin28A oncogene[53]. Lin28A recognizes a tetranucleotide sequence motif present in the terminal loop of certain precursors including let-7, miR-107, miR-143 and miR-200c, and recruits the terminal ribonucleotidyl transferase TUT4 (aka ZCCHC11) to add a poly-U tail to the 3´ end of the precursor to prevent Dicer from cleaving it.

1.2.2. IsomiRs with distinct regulatory functions

There are several examples where isomiRs expressed from the same miRNA gene have been described with distinct regulatory effects. The addition of nucleotides to the 3´ ends of mature miRNAs has been shown to impair target repression[54]. In 2009, Jones et al identified TUT4 as a key factor in miR-26 uridylation and found that uridylation decreased miR-26 mediated repression of one of its targets, interleukin-6[54]. Trimming variations in the 5´ end have also been linked to changes in target gene regulation[55]. Whilst 5´ variations are relatively rare compared to 3´ variations, the effect on target regulation is generally much more pronounced due to a shift in the miRNAs seed sequence, which has the potential to create an entirely new target pool. This effect was demonstrated in a study by Tan et

al who found that a 5´ variant of miR-9-1 gained the ability to downregulate two new targets, DNMT3B and NCAM2, compared to its canonical form, and also lost the ability to downregulate CDH1[55].

1.2.3. miRNA stability and turnover is regulated by 3´ tailing

By far the most common 3´ nucleotide additions to mature miRNAs involve Uridines and Adenosines, although limited evidence suggests Cytosines and Guanines may also be added after maturation [47,48,56]. Like pre-miRNAs, nucleotide additions are mediated by terminal ribonucleotidyl transferases such as TUT4, but also include other enzymes such as TUT2 (aka Gld2) and TUT7 (aka ZCCHC6)[56,57]. The TUTs vary in their preference for adding Adenosines or Uridines, with TUT2 preferring Adenosine as a substrate, and TUT4 and TUT7 preferring Uridine.

The regulatory consequences of nucleotide additions have been investigated in several studies, however our understanding remains limited and some of the studies have presented contradicting results. Most studies investigating uridylation in mature miRNAs support the idea that uridylation destabilizes miRNAs by serving as a marker for exonucleases such as DIS3L2. In an early study by Baccarini et al, they utilized an inducible pri-miR-223 vector in a cell line that does not express miR-223 in order to control miR-223 expression levels[58]. By inducing and then blocking expression of miR-223, they were able to track the miRNA during its decay. Small RNA sequencing revealed a decrease in overall miR-223 expression over time but a steady increase in the proportion of uridylated miR-223, which suggested uridylation played a role in miRNA turnover. Uridylation was also linked to a process called target directed miRNA degradation (TDMD), where high expression of a miRNA target can facilitate the degradation of the miRNA itself[59,60]. TDMD has been shown to have a greater effect on miRNA-target pairs with a high degree of complementarity[61]. It is believed that this high complementarity may partially release miRNAs from the RISC, exposing them to TUTs and exoribonucleases that lead to their degradation[61]. Adenylation has also been linked to TDMD and may occur under similar circumstances[60,62]. However, it has been suggested that most miRNA target sites do not possess the high degree of complementarity necessary to induce TDMD, and there are likely to be other mechanisms which contribute to nucleotide additions and effect miRNA function and stability[60].

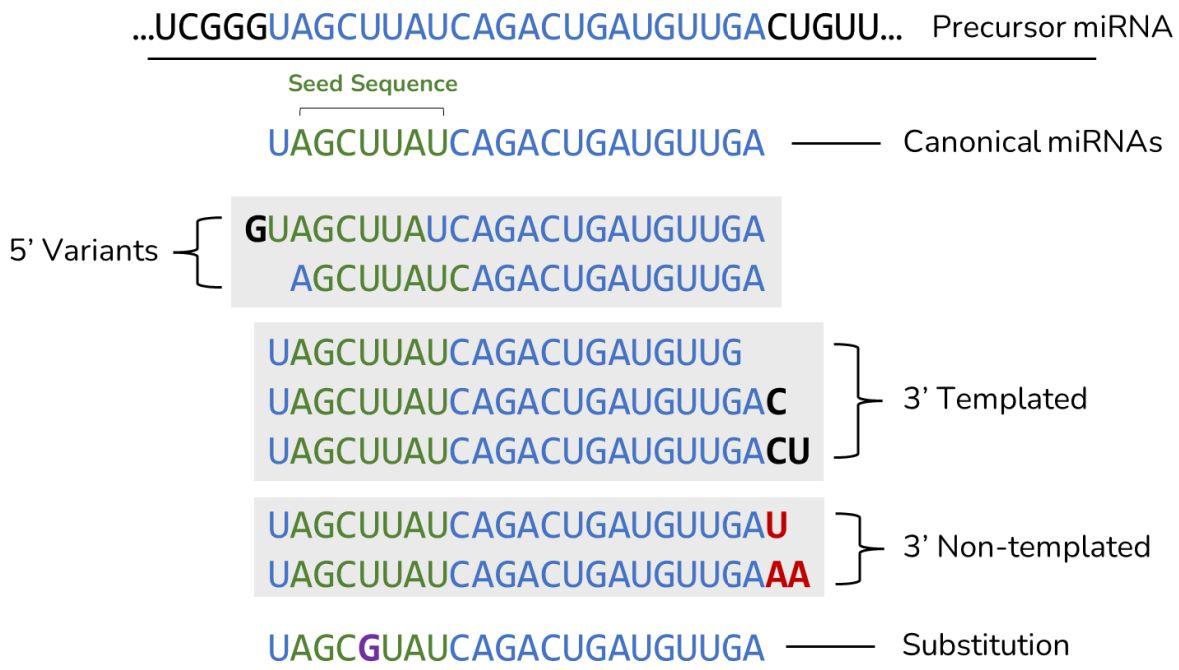

Figure 1.2. miRNAs are commonly expressed as various isoforms called isomiRs. Variants can have altered nucleotide sequences at the start (5′ variants), middle (substitution), or end of the sequence which are often distinguished by variants which match their precursor sequence (3′ templated) or not (3′ non-templated).

1.3. Argonaute is essential for miRNA-mediated translation repression

Mature miRNAs must form RNA-induced silencing complexes (RISCs) with one of the Argonaute proteins to participate in gene regulation[63]. Argonaute mediates the post-transcriptional regulatory activity of the complex by directly cleaving targets perfectly complementary to the miRNA (slicing), recruiting additional co-factors to facilitate degradation of transcripts, or interfering with their translation. The miRNA serves as a guide that directs the RISC to its gene targets via complementary base pairing to regions typically found in the 3´-untranslated region (3´UTR) of mRNAs. In humans this pairing is characteristically imperfect, which prevents direct cleaving of target mRNAs[64]. Therefore, alternative mechanisms such as translational inhibition and deadenylation are considered the primary means of target regulation[64]. The most important part of the miRNA which determines target recognition are the first 2-8 nucleotides from the 5´ end, named the seed sequence[65]. Although other parts of the miRNA are also known to influence miRNA-target pairing, many studies have shown that seed pairing is often sufficient[65]. The disparity between contributions of each part of the miRNA are a consequence of how they are bound to the Argonaute, as some parts of the miRNA are not exposed to the cytoplasm in a manner where they can contribute to base pairing with other RNAs[66].

In humans, there are four genes which encode Argonautes capable of binding miRNAs (AGO1, AGO2, AGO3 and AGO4) and inducing translational repression[67]. Argonaute 2 (Ago2), encoded by the gene AGO2, can directly cleave transcripts when loaded miRNAs are perfectly complementary to their targets[68]. Most of the functions for the four Argonaute proteins are believed to be redundant, which has been supported by studies utilizing knock-down experiments showing Argonautes typically compensate for each other[69]. One well known exception to this is during embryonic development, where Ago2's slicer activity is indispensable and knock-down leads to early developmental arrest[70]. More recently, studies have highlighted that Ago3 may also possess slicer activity but only when shorter (14nt) guide miRNAs are loaded[71].

1.3.1. Functional domains of Argonaute

All Argonautes have at least four highly conserved functional domains in common, named the PIWI/Argonaute/Zwille (PAZ) domain, P-element-induced whimpy tested (PIWI) domain, MID domain, and N-terminal domain (Figure 1.3)[67]. The PAZ, MID and PIWI domains all play a role in binding the miRNA. The MID and PIWI domains form a binding pocket with a metal-binding site, which anchors the phosphorylated 5' end of the guide strand during miRNA loading. Additionally, the PAZ domain binds the 3' end and is capable of recognizing the 3' overhangs of miRNA duplexes[72]. The N-terminal domain also plays a key role during miRNA loadings as it is critical for wedging the duplex to separate the guide and passenger strands. As mentioned previously, the PIWI domain contains the catalytic

domain responsible for the slicer activity of Ago2 and Ago3[71]. The PIWI domain's catalytic motif, harbouring a scissile phosphate, sits between the tenth and eleventh nucleotides of the miRNA and cleaves the bases of perfectly complementary target RNAs bound at this location[73].

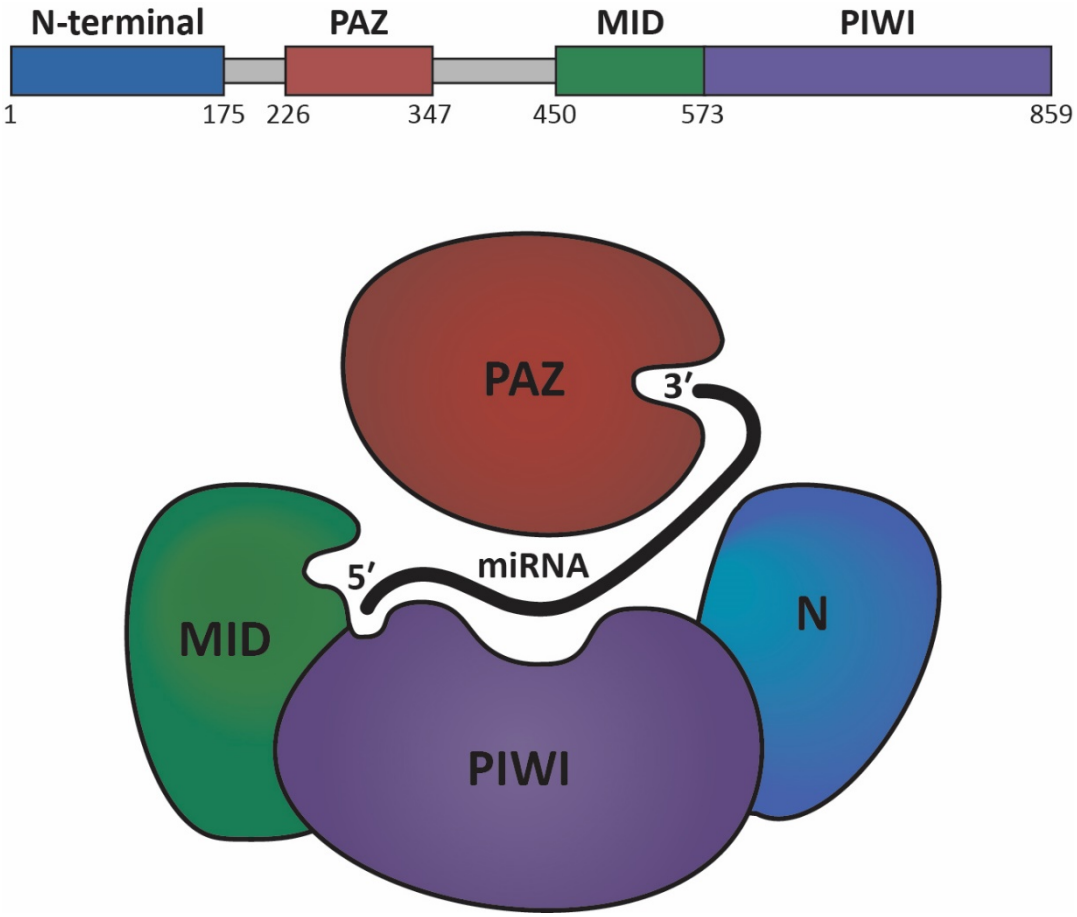

Figure 1.3. Argonaute consists of four primary functional domains – the N-terminal (N), PAZ, MID and PIWI domains. After RISC formation, the miRNA's 5′ phosphate is anchored in a pocket between the MID and PIWI domains while the 3′ end is bound to the PAZ domain. Numbers (top) show the positions of the start and end nucleotides for each domain.

1.3.2. Mechanisms of translational repression by the RISC

In mammals, RISCs are known to regulate gene expression through several distinct mechanisms such as the direct cleaving of transcripts, inhibiting translation, or transcript destabilization, which often operate in parallel (Figure 1.4A-C)[74]. Of these mechanisms, cleaving (Figure 1.4A) or translational inhibition (Figure 1.4B) is more rapid, although transcript destabilization (Figure 1.4C) is considered to have the largest influence on repression in animals[75]. The GW182 protein is a key adapter that binds to Argonaute and is crucial in both translational inhibition and transcript destabilization[76]. GW182 binds to Argonaute through its glycine-tryptophan (GW) repeat domain, and is recognized by tryptophan-binding pockets in Argonaute's PIWI domain[77].

Translational inhibition involves inhibition of the Cap-Binding complex eIF4F (Figure 1.4B) or displacement of Poly A Binding Protein (PABP)[74,78]. eIF4F is a complex comprised of eIF4E, eIF4G, and eIF4A, and is required for translation initiation. The RISC can interfere with translation initiation by preventing eIF4E from binding to the 7-methyl guanosine cap of mRNAs. During PABP displacement, GW182 interferes with PABP by binding to eIF4G and disrupting the loop structure formed during translation[74,78]. Furthermore, in Drosophila, GW182 was shown to interfere with eIF4A's association with the 5′ m^7G-cap on mRNAs, a requirement for translation initiation[79].

Like translational inhibition, destabilization of mRNAs also occurs via several mechanisms (Figure 1.4C). In addition to PABP, GW182 can also recruit two complexes – CCR4-NOT and PAN2-PAN3 – which are involved in a process called deadenylation[80]. During deadenylation, the poly(A) tail of the targeted mRNA is gradually removed, leading to dissociation of PABP, de-capping and finally degradation of the target RNA.

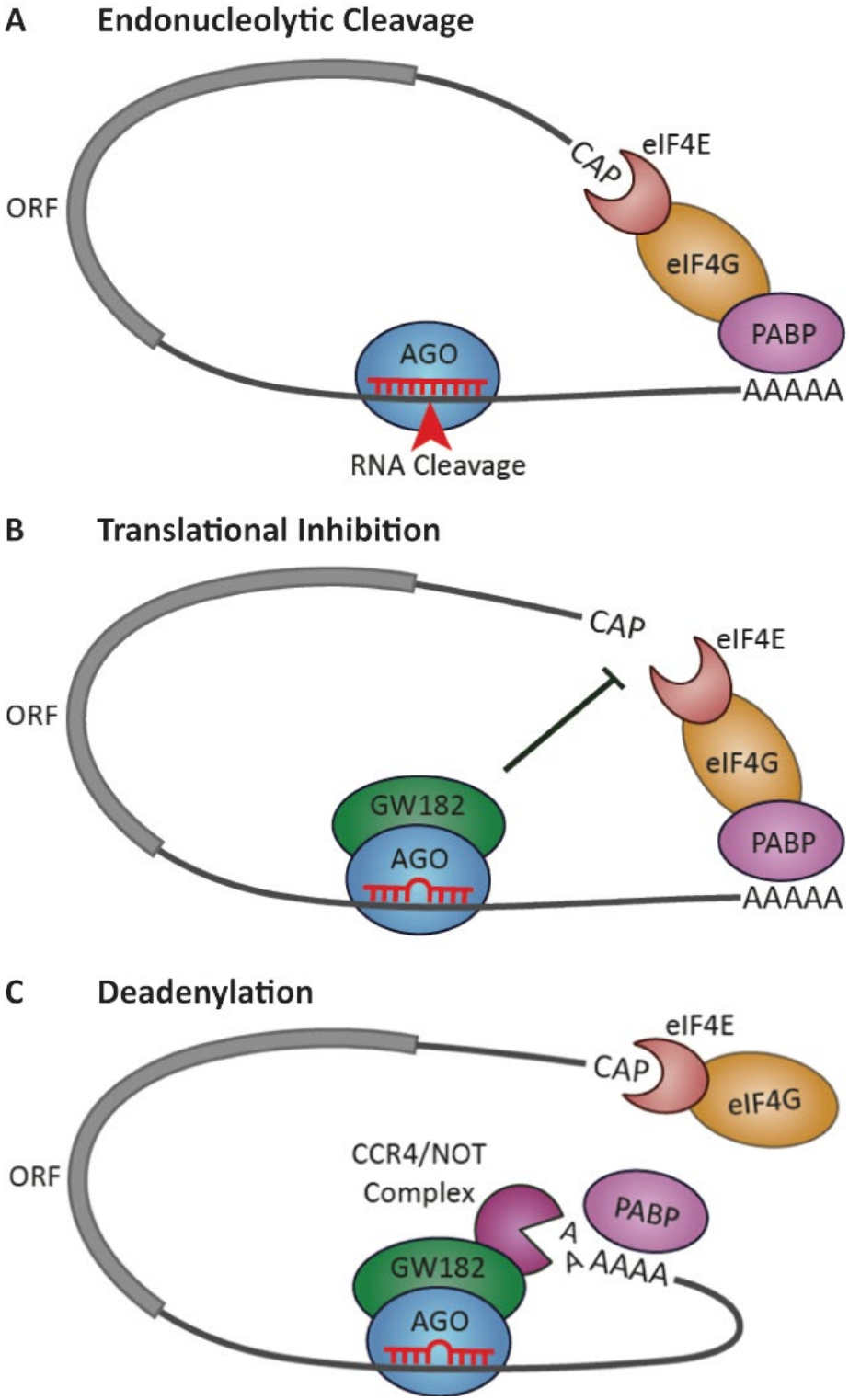

Figure 1.4. Multiple mechanisms facilitate post-transcriptional gene regulation with the RISC. **A.** Cleavage-competent Argonautes, such as Ago2, can directly cleave mRNA targets that are perfectly complementary to the miRNA. **B.** Translational inhibition depends on GW182, which can inhibit translational initiation. **C.** Destabilisation of mRNA targets through deadenylation is facilitated by GW182 and the CCR4-NOT complex.

1.4. miRNA dysregulation is a general feature in many cancers

Not long after the first human miRNA, *let-7*, was discovered in 2002 by the Ruvkun lab, miRNAs began to emerge as key participants in tumourigenesis[81]. In 2002, two miRNAs were identified as potential tumour suppressors due to their frequent down-regulation or deletion in chronic lymphocytic leukemia[82]. Calin et al later showed that many miRNA genes were located close to fragile sites or common breakpoints that frequently occurred in cancers, suggesting their loss of function was a key event[83]. Since then, expression studies have demonstrated that miRNAs possess a remarkable capacity to distinguish between different types of cancers or even sub-types and changes in miRNA expression often correspond with cancer prognosis[84–87]. In fact, miRNAs have been shown to play a role in all of the classical hallmarks of cancer – sustaining proliferative signalling, evading growth suppressors, invasion and metastasis, replicative immortality, angiogenesis, and apoptotic resistance[88–90].

A global downregulation of miRNAs is commonly observed in cancers, although individual miRNAs have been attributed a variety of oncogenic or tumour suppressive roles[91,92]. Evidence suggests that many of these miRNAs are important during tissue development and are vital for maintaining terminally differentiated states, with their misexpression underscoring a disturbance of the developmental process[93]. Additionally, cancer stem cells (CSCs), which retain many stem-like traits and represent a significant hurdle in therapeutic research, are also defined by distinct miRNA profiles when compared to their non-CSC counterparts[94].

miRNA expression or function can be altered as a direct consequence of impairments to the miRNA processing machinery, or from dysregulation caused by abnormalities of genetic or epigenetic origin, misexpression of transcription factors, or other dysregulated mechanisms which are involved in the same gene regulatory networks.

1.4.1. Impairments of the miRNA processing machinery in cancer

There are several mechanisms by which miRNA processing can contribute to cancer[95]. This includes changes in expression of individual biogenesis components, which can have downstream effects of increasing or decreasing miRNA expression, but also selective increases in particular miRNAs due to influences of co-factors which favour processing of certain miRNA species. Functional changes of processing machinery caused by mutations are also very common in some cancers, which typically leads to a loss of function.

The Microprocessor complex containing Drosha and DGCR8 is a point of vulnerability for tumourigenesis in many cells. In a recent systematic review, several single nucleotide polymorphisms (SNPs) in both genes were linked to cancer susceptibility[96]. In non-small cell lung cancer, overexpression of Drosha was identified and attributed to copy number gains and gene amplifications

which were present in a high proportion of cases[97]. DGCR8 has also been found upregulated in many cancers and has been linked to disease progression in both ovarian and prostate cancer[98,99]. DEAD box protein 5 (DDX5), another co-factor of the Microprocessor complex, can associate with SMAD proteins to drive increased processing of miR-21 primary transcripts into precursors[100]. SMADs are signal transducers of the transforming growth factor beta pathway and play a complex role in cancer progression[100]. Transport of the pre-miRNAs produced by the microprocessor complex can also be impaired by downregulation of the nuclear export protein exportin-5 (XPO5), either through reduced expression or phosphorylation by ERK kinase[101,102].

Clear links between dysfunctional miRNA processing and cancer development have been established in studies which revealed downregulation or deletion of DICER1 promoted tumour formation in mice[103]. Interestingly, in one study, this effect was exclusive to monoallelic but not biallelic loss of the gene, suggesting that DICER1 had an important tumour suppressive function but also had a role in malignant transformation[104]. Many studies have shown that specific DICER1 mutations could lead to an increased risk in developing cancers such as pleuropulmonary blastomas, rhabdomyosarcoma, cystic nephroma, and many endocrine tumours, although the functional consequences of many of these mutations has not been investigated thoroughly[105,106]. In lung cancer, DICER1 appears to have an oncogenic role as it is often amplified[97].

In cancers such as hepatocellular carcinoma, colon cancer, head and neck cancer and glioma, Ago2 is often overexpressed and has been linked to an increase in tumourigenesis, metastasis and patient prognosis[107,108]. However other studies have found Ago2 expression is reduced in cancers, such as in melanoma, which suggest that like many of the other biogenesis proteins, Ago2's role in cancer is dependent on the cell's miRNA profiles or expression of co-factors[109]. Another form of Ago2 dysregulation is through post-translational modifications, which have been shown to influence tumourigenicity. For example, Shen et al demonstrated that phosphorylation of Ago2 at Tyrosine 393 (Ago2-Y393), induced by hypoxic conditions, correlated with poorer overall survival in breast cancer patients[110]. In this study they were able to show this particular phosphorylation event was a consequence of interactions between epidermal growth factor receptor and Ago2, and that Ago2-Y393 increased cell survival and invasiveness by reducing its interactions with Dicer and subsequent miRNA loading[110]. In another study, acetylation of Ago2 at K720, K493, and K355 increased miR-19b levels by enhancing Ago2's capacity to recruit miR-19b precursors[111]. Ago2 acetylation at these sites was correlated with poorer prognosis in lung cancer patients. Interestingly, in contrast to many of the other miRNA biogenesis genes, mutations of Argonaute seem to be very rare[95].

Evidence suggests that some tissues may be more susceptible to forming cancers through dysregulation of general miRNA processing, as genome-wide association studies have found certain cancers have a much higher prevalence of mutations affecting the entire biogenesis pathway. Perhaps one of the strongest examples of a cancer where the miRNA processing machinery is likely to play a significant role in tumourigenesis is Wilms tumour. In 2014, a study by Torrezan et al[112] found mutations of miRNA processing genes in 33% of tumours, most commonly occurring in the Drosha gene but also with other mutations such as DICER1, XPO5, DGCR8 and TARBP2. These results were also supported by several other studies by Wu et al[113], Rakheja et al[114], Walz et al[115], Wegert et al[116], and Gadd et al[117]. In Rakheja et al's study they further examined the potential consequences of several of these mutations and found that Drosha mutations often led to a loss of RNAse IIIB activity which prevented processing of pri-miRNAs, leading to a global reduction in mature miRNAs. DICER1 mutations also frequently affected the RNAse IIIB domain, however this mutation only affected processing of 5p miRNAs from precursors, as DICER1 contains a second RNAse domain for 3p processing. As a result, this mutation led to a shift towards 3p miRNA maturation. These mutations have functional consequences for global miRNA expression and most likely favour expression of oncogenic miRNAs or reduce expression of miRNAs with tumour suppressive effects. In line with this, the let-7 family is predominantly 5p-derived and lower expression of several of its 5p members were found in both Drosha and DICER1 mutants in two of these studies[114,115].

1.4.2. Genetic aberrations of miRNA genes in cancer

As mentioned previously, many of the common genetic aberrations observed in cancers such as single nucleotide polymorphisms (SNPs), copy number variations and gene translocations, can affect the biogenesis genes involved in miRNA processing. However, genetic aberrations also occur in regions which harbor the miRNAs themselves or even in miRNA target sites and can cause significant disruptions to critical regulatory networks[92]. Interestingly, SNPs which alter the mature miRNA sequence are rarely reported and while some studies have found mutations to be common in primary miRNAs, these were shown to be inconsequential to the processing and expression of the mature miRNAs[118,119]. This contrasts with a number of cancer studies which have reported frequent occurrences of mutations in target sites that impact miRNA binding and subsequent regulation of their gene targets[120,121].

Copy number variations are pervasive in most cancers, and gains or losses of genomic regions containing miRNAs naturally lead to overexpression or reduction of miRNAs, respectively. Since Calin et al's initial observations highlighting the link between miRNAs and fragile sites, many more examples in cancer have been identified[82]. For example, loss of heterozygosity at region 14q32 is common in leukemia and corresponds with decreased expression of the miRNA cluster located there[122].

Additionally, in a study on gastric cancer, copy number gains of chromosomal region 8q24 were common and corresponded with overexpression of several miRNAs implicated in gastric cancer development[123]. Copy number gains of several miRNAs including miR-296, miR-324 and miR-3928, were identified in lung cancer and correlated with a poorer patient prognosis[124]. Finally, deletions or translocations of miR-15a and miR-16-1 were reported in a high number of chronic lymphocytic leukemia patients[82].

1.4.3. Epigenetic regulation of miRNAs is affected in cancer

Early studies established that in addition to protein coding genes, many miRNAs were also regulated by epigenetic mechanisms such as DNA methylation and histone modifications. One study found that nearly half of all miRNA genes had CpG sites located 2000 bp up or downstream[125]. In support of this, Saito et al had previously shown that treatment of cancer cell lines with demethylation agents or histone deacetylase inhibitors altered expression of miRNAs, with some miRNAs experiencing more than a 3-fold increase[126].

Both hypermethylation and hypomethylation have been linked to misexpression of miRNAs in cancer. Lujambo et al found extensive hypermethylation of miR-124a in several cancers, including colon, breast, lung carcinoma, leukemia and lymphoma and linked this hypermethylation to a loss of its expression in a colon cancer cell line[127]. A following study by the same authors showed that miR-148a, miR-34b/c and miR-9 were controlled by DNA methylation, and that silencing by methylation induced metastasis in cancer cells[128]. In papillary thyroid cancer, miR-146b expression was increased compared to normal tissue as a consequence of hypomethylation, and combined miRNA expression and methylation profiles could clearly distinguish malignant and benign tumours[129].

Enzymes involved in histone regulation, such as the polycomb-group proteins and histone deacetylases (HDACs), also cooperatively regulate miRNA expression alongside DNA methylation. Enhancer of zeste homolog 2 (EZH2), which forms part of the polycomb repressive complex 2 (PRC2), is upregulated in many cancers and is known to repress expression of a whole range of miRNAs, including miR-181a/b and miR-200a/b/c[130,131]. Overexpression of EZH2 and SUZ12, another PRC2 component, was also found to repress miR-31 and activate the NF-κB signalling pathway in leukemia[132]. Finally, in a study by Sampath et al, HDAC overexpression was linked to the epigenetic silencing of miR-15a, miR-16 and miR-29b in chronic lymphocytic leukemia, by showing an alleviation of expression upon HDAC inhibition in 35% of samples[133].

1.4.4. miRNAs are key mediators of transcription factor-driven oncogenesis

Dysregulation of transcription factors are a recognized feature in nearly all cancers[134]. Transcription factors and miRNAs often form feedback loops within regulatory circuits, and misexpression of either

can rapidly lead to disarray of gene signalling pathways and lead to oncogenesis or cancer progression (Figure 1.5). One of the earliest studies linking miRNA expression to an oncogenic transcription factor was by O'Donnell et al in 2005[135]. In this study they demonstrated that c-Myc could induce expression of the miR-17~92 cluster, and that several of these miRNAs could in turn regulate E2F1 transcription to control cell proliferation. In hepatocellular carcinoma, a negative feedback loop between c-Myc and miR-122 was identified, where c-Myc could repress miR-122 expression and miR-122 could prevent transcription of c-Myc through E2F1[136]. p53, another transcription factor and one of the most important tumour suppressors, promotes expression of miR-34 family, and many other miRNAs including miR-16-1, miR-143 and miR-145[137,138] . Many of p53's tumour suppressive functions such as cell cycle regulation and suppression of stemness traits and metastasis have been attributed to miRNA-mediated regulation of downstream gene targets[139].

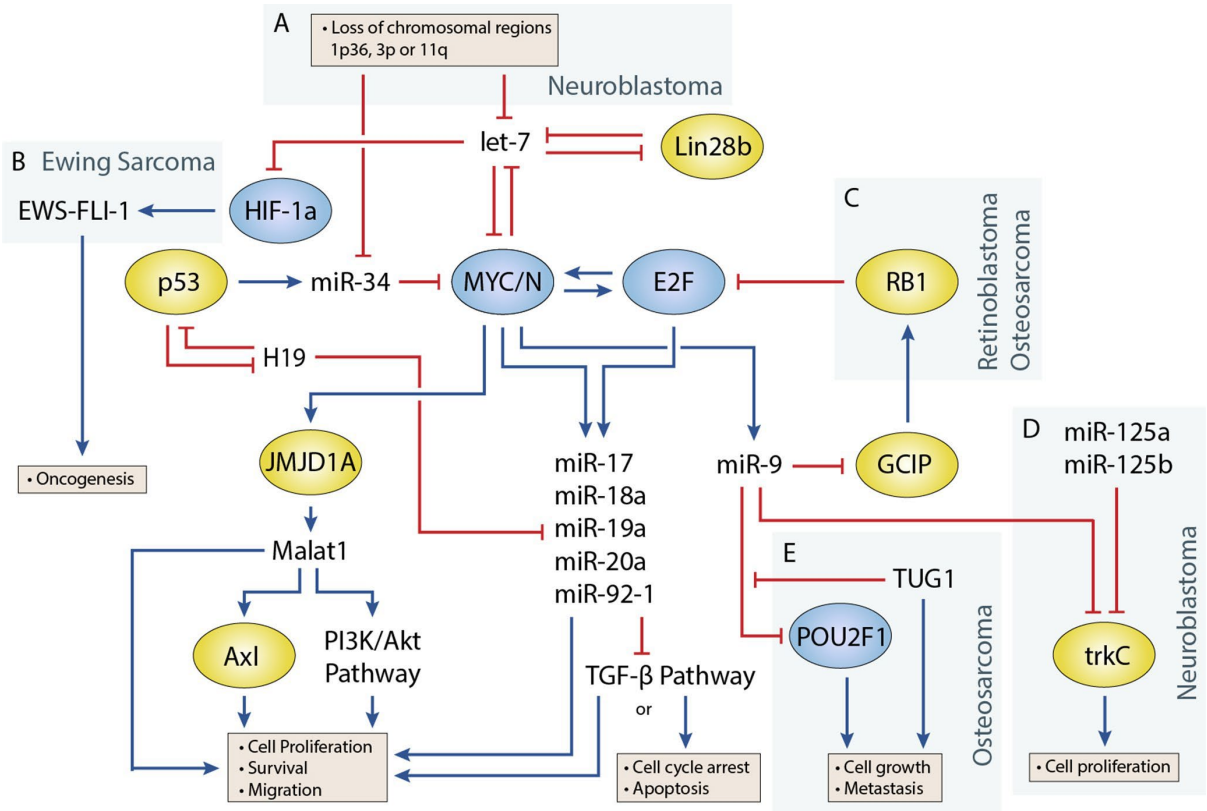

Figure 1.5. Gene regulation involves complex interactions between miRNAs, transcription factors, and long non-coding RNAs which all be dysregulated in cancers. Shows various elements of a key regulatory circuit involving MYC and E2F family transcription factors and their role in cancers. **A.** Loss of chromosomal regions where let-7 and miR-34 miRNAs are localized is frequently observed in cancers. **B.** let-7 dysregulation may facilitate overexpression of the oncogenic fusion transcript EWS-FLI-1 in Ewing sarcoma. **C.** The RB1 tumour suppressor regulates E2F, and loss of function can lead to the development of retinoblastoma. **D.** miR-9 can display tumour-suppressive properties by cooperating with miR-125a and miR-125b to suppress cell proliferation. **E.** The lncRNA TUG1 is suggested to act as a sponge, reducing miR-9 activity to supress tumour growth. Figure is from our published review in Frontiers in Genetics (See appendix A)[140].

1.4.5. IsomiR expression has been linked to cancer

Several studies have found isomiR expression is altered in cancers which suggests they may have a role to play the disease[3,141,142]. Telonis et al analysed 10,271 tumour datasets from The Cancer Genome Atlas and found that the presence or absence of isomiRs could effectively distinguish between 32 different cancers and that it could outperform classifications using miRNAs[142]. In another study looking at breast cancer, six 5´ isomiRs were found to be differentially expressed when comparing tumours to normal tissue[143]. One of these isomiRs, a 5´ variant of miR-140-3p, was more highly expressed, had novel gene targets, and possessed tumour suppressive effects not observed with its canonical counterpart. The authors proposed that the canonical miRNAs and isomiRs functioned synergistically as tumour suppressors. These studies have raised the possibility that isomiRs represent a previously ignored layer of gene regulation in cells and may have their own part to play in regulating oncogenesis.

Recently, isomiRs have been considered for their utility as blood/serum biomarkers for detecting or profiling cancers[144]. Circulating miRNA expression profiles have already been shown to be effective in distinguishing between diseased patients and healthy controls, however challenges remain in developing clinically useful biomarker assays[145,146]. One of the primary issues is the fact that miRNAs lack specificity and are broadly expressed in a variety of tissues in the body. Consequently, most miRNAs are present at some level in the blood regardless of a patient's condition and developing an unbiased assay which can distinguish between miRNA profiles from diseased and healthy patients may not be possible until the cancer has progressed significantly or is easily detected by other means. It is possible that isomiRs, which are potentially more unique in their expression, may offer a more effective means to identify cancers in their early stages when they are very difficult to detect.

1.4.6. Non-coding RNAs in pediatric solid tumours

Refer to appendix A for a review on miRNAs and long non-coding RNAs in pediatric solid tumours, written by myself and published in Frontiers in Genetics during my PhD[140].

1.5. Single cell sequencing reveals genetic and transcriptomic heterogeneity in cancers

Tumour heterogeneity describes the variation in molecular, cellular, or histological properties across tumours or tumour cells[147]. Tumour heterogeneity can either be inter-tumoural when this variation exists between different tumours of a cancer, or intra-tumoural when there is variability between the cells within a single tumour (Figure 1.6). Both inter and intra-tumour heterogeneity are common features in many different cancers including brain, lung, breast, colon and liver cancer, and can manifest as differences in genomic, epigenomic, transcriptomic, or proteomic profiles[148–152]. Critically, many of these differences detected on a molecular level have been shown to translate into clinically relevant phenotypical changes, such as altered tumour growth rates, metastatic potential, and drug resistance[153–155]. Consequently, this poses significant challenges for the development of accurate diagnostic or prognostic methods and has been cited as one of the primary reasons for drug failure in many cancers[156,157].

Research into inter-tumour heterogeneity has improved patient outcomes through targeted therapy approaches that tailor treatment plans based on cancer subtypes[152,158,159]. It is recognized that even with cancers of the same type, they may be driven by distinct biological pathways that are influenced by genomic or micro-environmental factors. Since specific biological pathways are often targeted in cancer therapy, this can have a significant impact on the efficacy of certain drugs. For example in breast cancer, the drug trastuzumab (Herceptin) is more effective against HER2-postive patients and is now routinely used in targeted therapy against this subtype, whereas Cyclin-dependent kinase (CDK) 4/6 inhibitors such as palbociclib, ribociclib, and abemaciclib are used for hormone receptor (HR)–positive and HER2-negative breast cancers[160,161]. While studies on tumour heterogeneity have led to improvements in drug efficacy for many cancers, this has mostly been through discoveries from bulk or population-level studies which target the same pathways across the whole tumour[155,159]. However, these pathways may not be suitable targets for all tumourigenic cells, as some are able to survive multiple rounds of treatments only to repopulate the tumour with more resistant cells[154,157]. It is increasingly recognized that a higher resolution is needed to study individual cells and their interactions within the tumour microenvironment in order to address all of the challenges that tumour heterogeneity poses[155,162].

At least two models have been described which explain the emergence of intra-tumour heterogeneity[155]. The clonal evolution model describes the emergence of genetically distinct lineages due to the inherent instability of the genome in cancer cells[158]. Over many generations, the proliferation of cancer cells leads to the divergence of cell populations with different genotypes, where those with selective advantages in tumour growth and survival become the most dominant populations. Because of the genetic diversity of cells, the likelihood of some cells surviving therapy

due to an inherent resistance is likely, and therapy may shift the selective pressure in favour of more resistant cell populations[158]. According to the cancer stem cell (CSC) model, some cancer cells adopt stem-like features, including cell proliferation and self-renewal, which are normally present in developing stem cells[163,164]. Cancer stem cells can produce more differentiated progeny with distinct phenotypes to promote tumour growth and survival, and evidence suggests differentiation of these cells is often incomplete and reversible under certain circumstances[165]. Both models have supporting evidence in many cancers and are likely to act together in producing the heterogeneity observed across genetic, epigenetic, transcriptomic, and proteomic layers[166]. Proper characterization of tumour subpopulations with clinical relevance has remained extremely challenging in the past few decades[162]. However, recent developments in cell isolation and sequencing methods have offered researchers new tools to study the tumour microenvironment and there is renewed hope that this will lead to significant improvements in the outcomes of cancer patients.

1.5.1. Single cell sequencing has enabled deeper interrogation of intra-tumour heterogeneity

The introduction of next-generation sequencing (NGS) ushered a new era in oncology by enabling the comprehensive profiling of cancer genomes and transcriptomes[167]. Their adaptation to single cell technologies has been facilitated by advances in automation and microfluidics, which can capture or isolate thousands of individual cells and generate cost-effective sequencing libraries all within a single experiment[168]. Additionally, the constraints imposed from having to sequence thousands of separate libraries have become much less of an issue due to the remarkable drop in sequencing costs seen over the past few decades. In 2007, the cost of sequencing a single human genome was approximately $10 million US dollars whereas in 2022 this can be done for only $600[169,170]. Currently single cell sequencing encompasses a burgeoning field of NGS tools, including multi-omic methods which can simultaneously measure information across different modalities, and are revealing critical details about tumour cells and their microenvironment[171].

There are a number of single cell DNA sequencing methods available for whole genome sequencing, exome sequencing or targeted gene sequencing, used in oncology research for determining genomic heterogeneity, evolutionary trajectories of subclones, and detection of rare mutations[172]. Some of the earliest single cell studies into clonal evolution established a punctuated model for copy number evolution, demonstrating that copy number alterations were distinct events occurring early in oncogenesis, whereas single nucleotide variations were more gradual events throughout the course of the disease[150,173]. Furthermore, the genetic diversity of tumours is beginning to be appreciated as many subclones, including some with drug resistant genotypes, were found to exist in small cell numbers prior to treatment[174–176]. Recent work has highlighted some of the limitations of traditional

therapies developed from population level studies and indicates that routine biopsies are unlikely to capture the full range of clinically relevant genomic mutations[177].

Of all the single cell sequencing methods, single cell RNA sequencing is the most well developed and widely adopted technology[168,178]. Single cell RNA sequencing allows quantification of gene expression through the detection of RNA transcripts. Methods of single cell RNA sequencing are often divided into those which utilize traditional cell sorting with microwell plates and those using droplet microfluidic devices. The traditional plate-based assays, such as Smart-seq and CEL-seq, were some of the earliest methods to be developed for full transcriptome analysis[179,180]. Droplet-based protocols such as inDrop and Drop-seq were subsequently developed and were a major milestone in single cell sequencing as they made it feasible to sequence many thousands of cells at once, although were restricted to gene level quantification as full-length capture was not possible[181,182]. The power of single cell RNA sequencing has been demonstrated across many cancers where it has been used not only to characterize cell populations, but also to map out their differentiation trajectories, predict copy number variations and identify gene modules or networks that may predict their clinical presentation[148,183]. One of the major insights derived from single cell RNA sequencing is the observation that cell subpopulations derived from transcriptomic analysis are often decoupled from any apparent genetic heterogeneity. In oligodendroglioma, heterogenous subpopulations bearing resemblance to developmental pathways were all present across different genetic subclones[183]. In another brain cancer, glioblastoma, the authors transplanted genetic subclones into patient-derived xenografts and found that the full range of transcriptomic cell populations, including their relative abundances, were recapitulated in the new tumours[184]. Collectively, this work suggests that epigenetic or post-transcriptional gene regulatory mechanisms may be the predominant driving force of this form of heterogeneity[151].

Furthermore, the number of available single cell sequencing methods for detecting alternative molecular profiles continues to grow, and it is now possible to measure epigenetic profiles such as DNA methylation (scRRBS), chromatin accessibility (scATAC-seq), and histone markers (scChIP-seq), and other classes of RNAs not captured in standard RNA-seq such as small RNAs (Small-seq)[185–188]. Multi-omics approaches which combine multiple measurements are also available, although the technical challenges associated with sequencing each type of information is often compounded and have not yet been adopted as widely[171,189–192].

1.5.2. The current state of single cell small RNA sequencing

miRNAs are involved in the regulation of nearly all biological processes and are therefore likely to play an important role in the post-transcriptional mechanisms that drive differentiation and cell state

regulation. However, sequencing technologies which can capture this class of small RNAs and provide sufficient resolution to interrogate them on a single cell level remain underdeveloped. Several methods have been recently developed specifically for single cell small RNA sequencing, generally utilizing a two-adapter ligation approach for small RNA capture and library preparation[188,191,193–195]. With two-adapter ligation, a 5' pre-adenylated adapter is ligated to the 3' end of RNAs after cell lysis, followed by ligation of a second adapter to the phosphorylated 5' ends of small RNAs. cDNA is generated through reverse transcription using a primer which is reverse complementary to the adapter ligated at the 3' end and then subsequent rounds of polymerase chain reaction (PCR) amplify the cDNA for sequencing. Relative to many other single cell sequencing methods, small RNA sequencing is still technically challenging, expensive, and has limited cell throughput as they are all restricted to plate-based assays. The first of these methods was the Small-seq protocol developed in 2016, which was later followed by a CleanTag adapter-based protocol[188,193]. Two multi-omic methods were subsequently developed, Holo-seq and a microRNA-mRNA co-sequencing protocol by Wang et al, capable of sequencing small RNAs and mRNAs from the same cell[191,194]. A key distinction in the Holo-seq method was the use of removable carrier RNAs, introduced after cell lysis, to protect RNAs from degradation during library preparation. To build separate small RNA and mRNA libraries, poly adenylated RNAs were isolated from the cell lysate using targeted magnetic beads in the Holo-seq protocol, whereas in Wang et al's co-sequencing protocol cell lysates were split in half. Currently, research into the role of the miRNA pathway at the single cell level and its contribution to tumour heterogeneity has been very limited. In the Holo-seq study, 32 hepatocellular carcinoma cells were sequenced for their mRNA and miRNA profiles[194]. Interestingly, this work highlighted 3 distinct populations of cells when analysing their mRNA profiles and only 2 when analysing the miRNA profiles. The miRNA-based populations corresponded closely with 2 of the mRNA populations, indicating that miRNAs can reveal cell subpopulations with distinct gene expression profiles. Further research is necessary to understand whether heterogenous miRNA expression is a common feature in cancers and if the miRNA pathway contributes to heterogeneous gene regulation in cancer subpopulations.

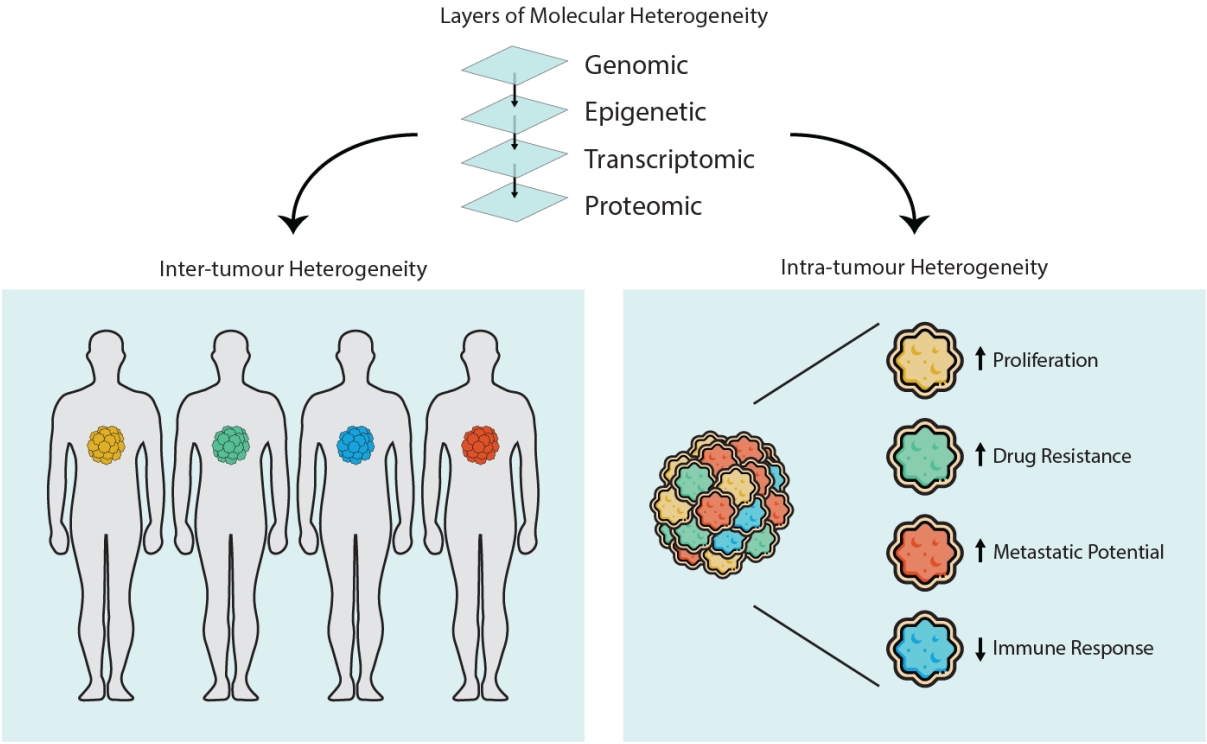

Figure 1.6. Molecular heterogeneity is evident in genomic, epigenetic, transcriptomic, and proteomic profiles. In cancer, heterogeneity often exists between patients (inter-tumour heterogeneity), which can impact prognosis. Additionally, cancer cells can take many forms within a single tumour and adopt a variety of phenotypes which support the tumour microenvironment. Both forms of heterogeneity challenge our ability to develop effective therapies.

1.6. Summary and research questions

miRNAs play an essential role in gene regulation through suppression of translation and destabilization of messenger RNAs[196]. miRNAs are one of the most abundant regulatory molecules in human cells and are heavily integrated into gene networks driving essentially all biological processes, including regulation of cell differentiation, the cell cycle, and intercellular signalling[197,198]. Their dysregulation is a common feature in many diseases, including cancer, and studies have demonstrated functional roles for miRNAs in oncogenesis and drug resistance[92]. Regulation of miRNA biogenesis is incredibly complex, involving a multitude of proteins that work together to process primary miRNA transcripts into their mature, biologically active forms[1].

miRNA processing during biogenesis, and after maturation, produces a variety of miRNA isoforms called isomiRs[199]. Research shows that at least some of these isomiRs are functional and can either act together to suppress the same targets or completely different ones[55]. IsomiR expression is strongly associated with different types of cancer, however our understanding of their contribution to oncogenesis remains limited[142,200]. Furthermore, the mechanisms behind isomiR processing are still poorly understood and it is not known how precise isomiR expression can be regulated cell autonomously. Investigating this may provide insight into their contribution to the heterogeneity observed in cancer cells[154]. Prior to this thesis and to the best of our knowledge, no studies have been published by others which investigate cell-wide isomiR expression and processing at the single cell level.

In the final stage of miRNA biogenesis, the RNA-induced silencing complex is formed, which contains two essential components for miRNA regulation – the mature miRNA and the protein Argonaute[67]. RNA sequencing studies have detected alternatively spliced transcripts of Argonaute with protein-coding potential[68,201]. However, it is not known how these isoforms affect the function of Argonaute and its interaction with miRNAs in humans.

This thesis is composed of two hypotheses which aim to address major gaps in our understanding of the miRNA pathway. The first investigates the miRNA pathway's role in intra-tumour heterogeneity, using single cell sequencing data to relate miRNA expression to cell states and to understand isomiR regulation in individual cells. The second investigates the biological function of isoforms of Argonaute 2.

1.6.1. Hypothesis 1 - microRNA and isomiR expression contributes to heterogeneous gene regulation and can identify subpopulations in cancer

- **Aim 1**: Identify miRNAs that are differentially expressed in glioblastoma cells and investigate their association with known glioblastoma cell states
- **Aim 2**: Characterize isomiR expression on a single cell level and assess their potential to regulate processes involved in intra-tumour heterogeneity

For Hypothesis 1, the first aim is addressed in chapter 2 and the second aim in chapter 3 in this thesis. Chapter 3 is an extended version of our published paper in Scientific Reports[202].

1.6.2. Hypothesis 2 - Splice variants of Argonaute 2 have distinct regulatory functions through association with different RNAs

- **Aim 1**: Quantify Ago2 isoforms in human tissues and cell lines to identify isoforms with potential biological significance
- **Aim 2**: Compare the small RNA and miRNA binding profiles of Ago2 isoforms

For Hypothesis 2, both aims are addressed in chapter 4.

2. The role of miRNAs in glioblastoma tumour heterogeneity and cell state regulation

2.1. Chapter introduction

2.1.1. Heterogeneity is a hallmark feature of the brain cancer glioblastoma

Glioblastoma is one of the most common and aggressive gliomas, which originates in the brain, and has seen few improvements in patient outcomes over the past few decades[203]. One of the main reasons for this lack of progress is that extensive tumour heterogeneity is one the hallmark features of this disease[204]. Inter-tumour heterogeneity between tumours of different patients, as well as intra-tumour heterogeneity of cells within a tumour, challenge the development of drugs that can effectively treat glioblastomas with complex molecular profiles and varying phenotypes[204,205].

Recent technological advancements have paved the way for molecular subtyping using next generation sequencing. In glioblastoma, inter-tumour heterogeneity has been characterized on genetic and transcriptional levels. In 2010, Verhaark et al integrated gene expression data from The Cancer Genome Atlas (TCGA) and identified four subtypes of glioblastoma – which they named proneural (TCGA-PN), neural (TCGA-NE), classical (TCGA-CL), and mesenchymal (TCGA-MS)[206]. The subtypes were later revised to exclude the neural subtype, whose classification was attributed to contamination with normal tissue[207]. Glioblastoma subtypes are strongly associated with genomic events of specific genes. For example, classical subtype tumours were found to be more frequently affected by EGFR amplifications, aberrations in PDGFRA and IDH1 more common in proneural subtypes, and loss of NF1 common in mesenchymal subtypes[206,207]. As overall survival times are different in tumours from each of these TCGA subtypes, genes associated with these subtypes are considered to have clinical value as prognostic biomarkers[207,208].

2.1.2. Evolution of models for intra-tumour heterogeneity in glioblastoma

One of the first discoveries of cancer stem cells was in brain tumours, which formed the basis of early models for intra-tumour heterogeneity in glioblastoma[209]. Purportedly, glioma stem cells existed in small numbers but were self-renewing and could generate the diverse cell populations that tumours were comprised of[203]. Initial studies suggested only cell populations enriched in glioblastoma stem cell markers (e.g CD113) could generate tumours[209]. However, later studies identified multiple cell populations with stem-like or tumourigenic properties, isolated using different cell markers indicating a far more complex hierarchy[210,211].

Models for glioblastoma intra-tumour heterogeneity have expanded significantly since the development of high throughput single cell sequencing[184,212,213]. One of the earliest single cell RNA

sequencing (scRNA-seq) studies in glioblastoma identified common sets of genes, or meta-signatures, expressed across 5 different patient tumours related to hypoxia, immune response, oligodendrocytes, and the cell cycle[212]. In a study by Neftel et al, gene expression data was incorporated from 28 glioblastoma tumours using scRNA-seq and 401 tumours using RNA-seq from TCGA, to construct a more unified model of glioblastoma[184]. In this model they described four main cancer subpopulations, which they referred to as cell states, resembling various cell types that exist during normal brain development[184]. This included astrocyte-like (AC), neural-progenitor-like (NPC), oligodendrocyte-progenitor-like (OPC), and mesenchymal-like (MES) cell states. Interestingly they found that all tumours were comprised of multiple cell states, although not necessarily containing all states at once. The presence and relative frequencies of each cell state were shown to directly influence TCGA subtype classifications, with TCGA-CL tumours more dominant with cells in an AC cell state, TCGA-PN tumours more dominant in NPC and OPC cell states, and TCGA-MS tumours with MES states. Furthermore, plasticity was clearly demonstrated between each cell state, which could all initiate tumours that eventually recapitulated the relative frequencies of cell states observed in the original tumours[184]. An additional cell population resembling outer radial glia was identified in a later study by Bhaduri et al, which contributes to the invasiveness of glioblastoma[214].

2.1.3. miRNAs play a key role in brain development and glioblastoma

miRNAs play an essential role throughout the developing brain, contributing to cell fate specification and differentiation in many neural or glial stem/progenitor cells[215–217]. Extensive documentation of miRNA dysregulation in glioblastoma and their effects on key cancer pathways suggests they are also important in tumourigenesis[218–224]. Furthermore, miRNA expression profiles can significantly improve classification of tumours with TCGA subtypes, making them potential biomarkers and also suggesting their activity is intrinsically linked to the gene networks that drive each of these subtypes[225]. As research has now highlighted a direct link between TCGA subtypes and cell state compositions on a single cell level, this implies that miRNA expression is also associated with these cell states and may have important functions in cell state regulation[184]. Despite this, there is an absence of studies in the literature which aim to investigate miRNA expression in glioblastoma single cells and the role miRNAs play in the regulation of cell states. We hypothesized that cell states should be distinguishable by analysis of miRNA expression and by examining glioblastoma cells we may find evidence of autonomously regulated miRNAs of importance to the cell states described in Neftel et al's glioblastoma model[184].

2.2. Specific methodology

2.2.1. Data collection

Raw single cell small RNA sequencing reads were obtained from the Gene Expression Omnibus (GEO) database under accession id GSE81287.

Raw or normalized gene counts and metadata from glioblastoma single cell RNA sequencing was obtained the Gene Expression Omnibus (GEO) database under accession ids GSE84465 (Darmanis dataset), GSE131928 (Neftel and Neftel_10X datasets), GSE102130 (Filbin dataset), and GSE57872 (Patel dataset)[184,212,213,226].

Read counts from bulk miRNA and RNA sequencing of glioblastomas were obtained from the GDC Data Portal under project ID CPTAC-3[227]. Files for miRNA isoform expression were used to quantify mature miRNAs.

2.2.2. Single cell small RNA sequencing pre-processing and mapping

For the single cell small RNA sequencing reads, UMI sequences were removed prior to any adapter removal and appended to the read headers. Adapters were then removed using cutadapt (v2.7) with a minimum overlap of 1 nt and maximum error rate of 0.1 between reads and adapter sequences[228]. After UMI and adapter removal, reads shorter than 15 nucleotides were excluded. To identify duplicated reads, reads were aligned to the human genome (hg38) using bowtie (v1.2.3) with the following parameters: -n 2 -e 120 -l 20 --best[229]. Human-aligned reads were subsequently deduplicated with umitools (v1.0.0) with default settings[230].

Processed reads were aligned to miRbase (v22.1) annotated precursor miRNAs using miraligner (v3.4), with the following parameters: -sub 1 -trim 3 -add 3[36,231]. Reads which successfully aligned to a miRNA were also annotated with any variants to the miRbase defined mature sequence and converted to miRNA and isomiR count matrices.

2.2.3. Identification of cell subpopulations with miRNA expression

To identify glioblastoma cell subpopulations using miRNA or isomiR expression, miRNA and isomiR count matrices were analysed with the Seurat R package (v3.1.5)[232]. miRNA or isomiR features present in 3 or less cells, or cells with less than 1000 mapped reads, were excluded from analysis. For dimensionality reduction, we used PCA on the top 25% of variable features after centering and scaling the data. Cells were visualized using Uniform Manifold Approximation and Projection (UMAP) with a range of principal components and resolutions, determined with JackStraw plots. This was followed by cell clustering using Seurat's inbuilt graph-based clustering method. Clustering of cells was stable across a range of parameters and were deemed acceptable due to the presence of cluster specific

miRNA 'markers' highly expressed in one population and absent in the other. For differential expression analysis, only features with an adjusted p-value (p.adj) less than 0.05 and $\log_e$ fold-change larger than 1 were considered differentially expressed.

2.2.4. miRNA target prediction

miRNAtap (v1.28) package was used to aggregate target predictions across five different algorithms (DIANA, Miranda, PicTar, TargetScan, and miRDB) using the 'minimum' method and only considering targets predicted by 2 or more algorithms[233].

2.2.5. Scoring for TCGA subtypes, cell states and miRNA targets

Scores which reflected expression of a set of genes (i.e gene modules) were calculated using a method similar to that described in the Neftel et al study for single-cell gene signature scores[184]. First, gene counts for all samples were converted to $\log_2$ transcripts-per-million ($\log_2$[TPM+1]) and centered by deducting expression of each gene by its mean expression across all samples from the same dataset. Then for each sample, scores were calculated from the mean expression of the gene module minus the mean expression of a control gene set. For the control gene set, aggregate expression of each gene across all samples from the same dataset were used to sort and separate genes into 30 expression bins. For each gene in the gene module, 100 genes were randomly selected from the same expression bin and placed in the control gene set.

The gene sets for the 3 TCGA subtypes were obtained from Wang et al's study[207]. Cell state gene sets were obtained from Neftel et al's study[184]. Gene sets for miRNA targets were generated by miRNAtap[233].

2.2.6. Code for data analysis and figures

Documents containing code used to generate the results in this section can be found in the following link: https://cloudstor.aarnet.edu.au/plus/s/GaNTw7rMWztjliu

2.3. Identification of heterogeneously expressed miRNAs in glioblastoma

2.3.1. Distinct glioblastoma subpopulations can be identified through single cell miRNA and isomiR expression profiles

Previous work by Neftel et al, using unsupervised hierarchical clustering with scRNA-seq expression data, identified 6 gene expression modules which were associated with 4 different cell states in glioblastoma[184]. These gene expression modules spanned multiple glioblastoma tumours, suggesting they represented commonly occurring cell states in glioblastoma. Their work also showed that glioblastoma cell states emerge spontaneously and are predominantly driven by cell autonomous factors including genomic perturbations[184]. We hypothesized that unsupervised clustering of miRNA expression from glioblastoma cells may also identify subpopulations in different cell states. Additionally, we investigated isomiR expression to explore their potential role in cell state regulation. To determine if miRNA or isomiR heterogeneity may contribute to cell state regulation or identity in glioblastoma, we analysed 173 cells from 3 glioblastoma primary cultures and 1 glioblastoma cell line[232]. We used Seurat's built-in graph-based clustering approach on UMAP projections for both miRNA (Figure 2.1A) and isomiR (Figure 2.1B) expression and found most cells co-localized with other cells from their respective cell type with minimal evidence of common miRNA or isomiR expression modules across different cell types[232]. However, we observed that one of the glioblastoma primary cultures, named KS4, formed two distinct groups that were identifiable in both the miRNA (Figure 2.1C) and isomiR (Figure 2.1D) data, which we considered may represent two cell subpopulations in different cell states. The two groups of cells formed using miRNA and isomiR expression were nearly identical, except for two cells (Figure 2.1E).

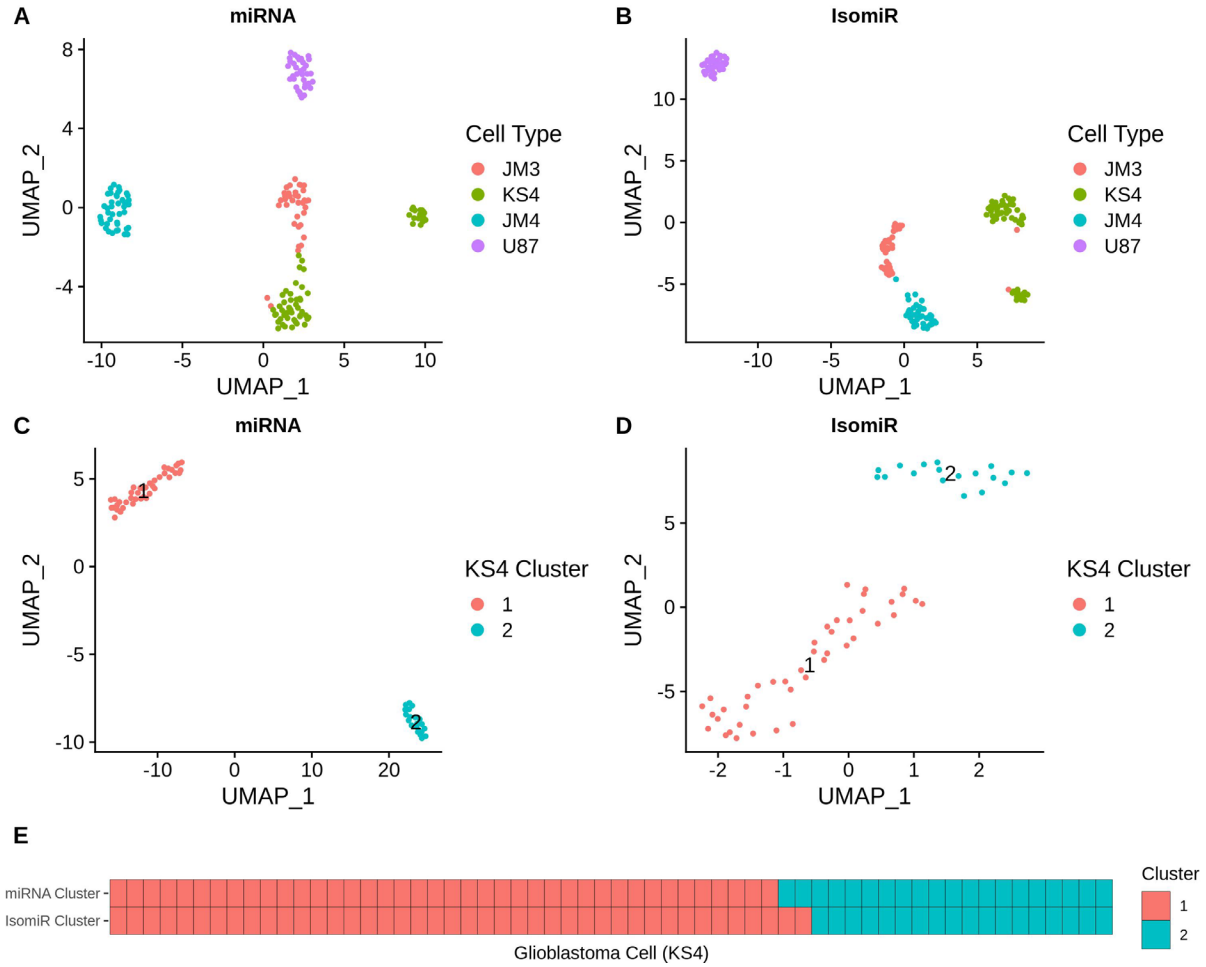

Figure 2.1. Identification of heterogenous glioblastoma cell subpopulations from miRNA and isomiR expression. UMAP plots (top) containing all four glioblastoma cell types using **A.** miRNA and **B.** isomiR expression. UMAP plots (bottom) containing cells from the KS4 glioblastoma cell type for **C.** miRNA and **D.** isomiR expression, coloured by their assigned clusters using Seurat's in-built clustering algorithm[232]. **E.** miRNA and isomiR clustering results for each KS4 cell.

2.3.2. Expression of two miRNA clusters distinguish KS4 glioblastoma subpopulations

Following this, we compared miRNA expression between the two KS4 cell clusters using differential expression analysis (Figure 2.2). In total, we identified 20 miRNAs which were significantly upregulated in KS4 cluster 1, and 5 miRNAs which were significantly upregulated in KS4 cluster 2 ($p.adj < 0.05$). Further investigation into the upregulated miRNAs in KS4 miRNA cluster 1 found that 18 of the 20 upregulated miRNAs originated from a miRNA cluster hosted within the DLK1-DIO3 gene locus on chromosome 14q32 (Figure 2.2B and Figure 2.3). We also found that 3 of the 5 upregulated miRNAs in cluster 2 - miR-224-5p, miR-224-3p, miR-452-5p, were also from a single miRNA cluster on chromosome X (Figure 2.2C). By comparing the expression of the miRNAs significantly upregulated in cluster 1 or 2 (Figure 2.4), we found that the miRNAs from each of these miRNA clusters dominated expression (>75%). Together this data shows that some miRNAs, including two miRNA clusters, can be expressed cell-autonomously and are potential regulators of cell states on a single cell level.

We also compared the expression of isomiRs in the KS4 clusters (Figure 2.5). We considered two possible scenarios which would lead to differentially expressed isomiRs – that isomiR processing was altered in these two cell clusters or that the isomiRs were reflecting differences in miRNA expression. If isomiR processing was altered, then we would expect to find differentially expressed isomiRs which are not in proportion to the expression of other isomiRs from the same miRNA. Differential expression analysis identified 5 isomiRs upregulated in cluster 1 and 13 isomiRs upregulated in cluster 2. However, we found that many of the isomiRs upregulated in either KS4 cluster matched the miRNAs identified previously, and that the isomiRs were generally in proportion to expression of the other isomiRs from the same miRNA gene (Figure 2.5B-C), suggesting miRNA expression, not isomiR processing, was the key differentiating factor between these two cell populations.

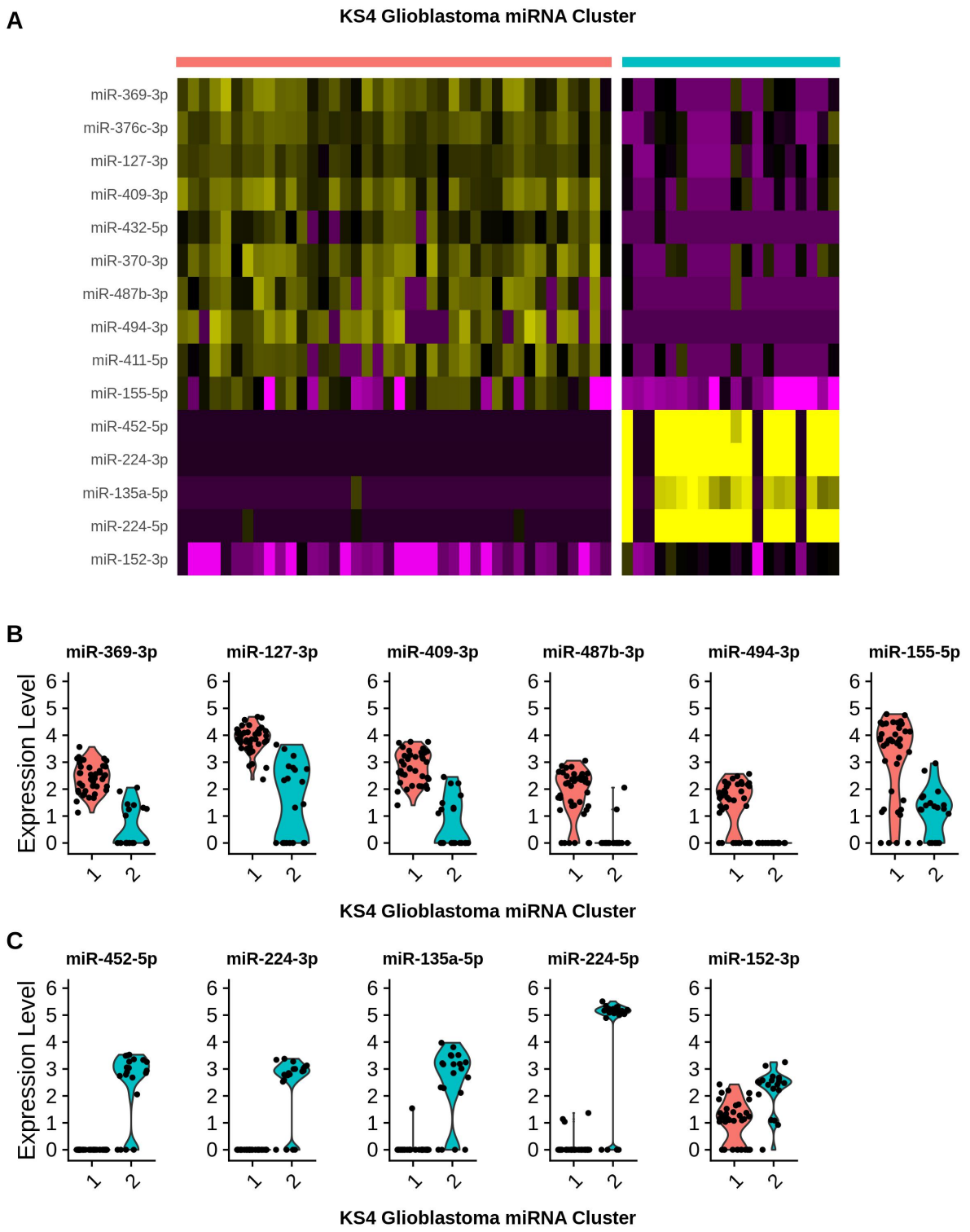

Figure 2.2. Comparison of miRNAs between KS4 glioblastoma cells in miRNA-based clusters 1 (red) and 2 (blue). **A.** Heatmap containing miRNAs with the highest fold change expression between clusters. Not all miRNAs shown here are statistically significant. **B.** Top 6 miRNAs by statistical significance that are differentially upregulated in KS4 cluster 1 (p.adj < 0.05). **C.** The 5 miRNAs differentially upregulated in KS4 cluster 2 (p.adj < 0.05). miRNA expression shown is in $\log_e$.

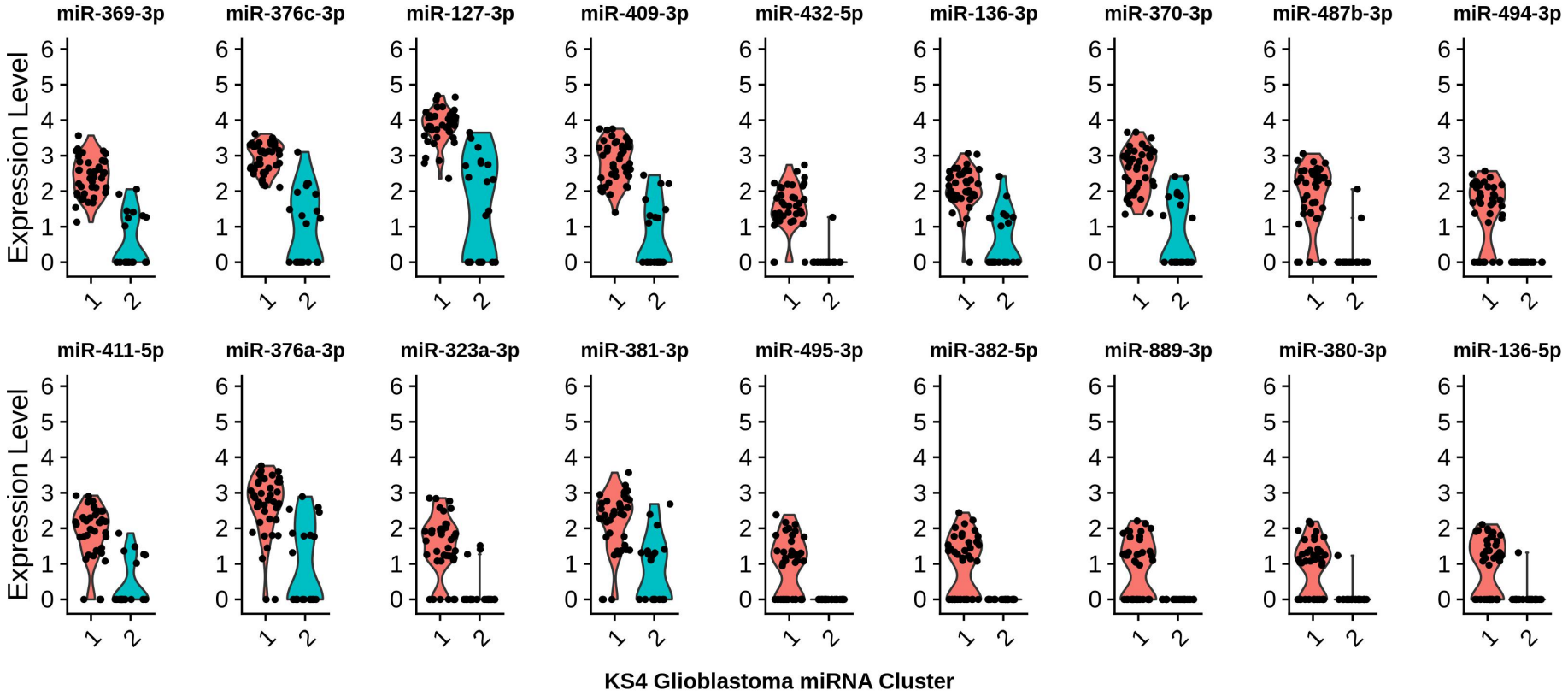

Figure 2.3. DLK1-DIO3 locus derived miRNAs upregulated in KS4 Glioblastoma miRNA-based clusters 1 (red) and 2 (blue). Only miRNAs with adjusted p-values less than 0.05 were included. Expression shown is in $\log_e$.

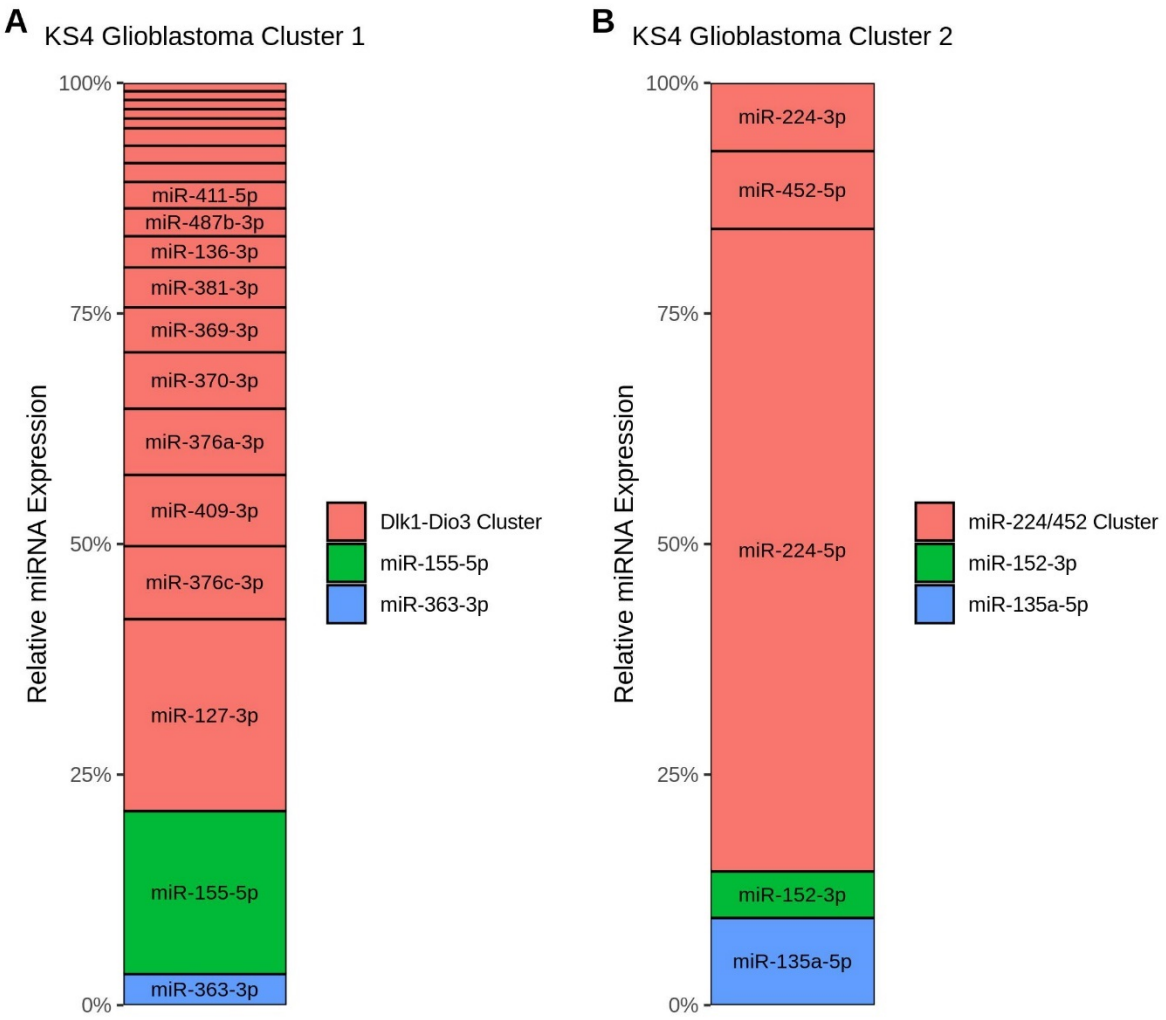

Figure 2.4. Expression of upregulated miRNAs in two KS4 cell subpopulations is dominated by two miRNA clusters. **A.** Shows relative expression of the 20 miRNAs upregulated in glioblastoma cells from KS4 Cluster 1 (compared to KS4 Cluster 2). **B.** Shows relative expression of the 5 miRNAs upregulated in glioblastoma cells from KS4 Cluster 2 (compared to KS4 Cluster 1).

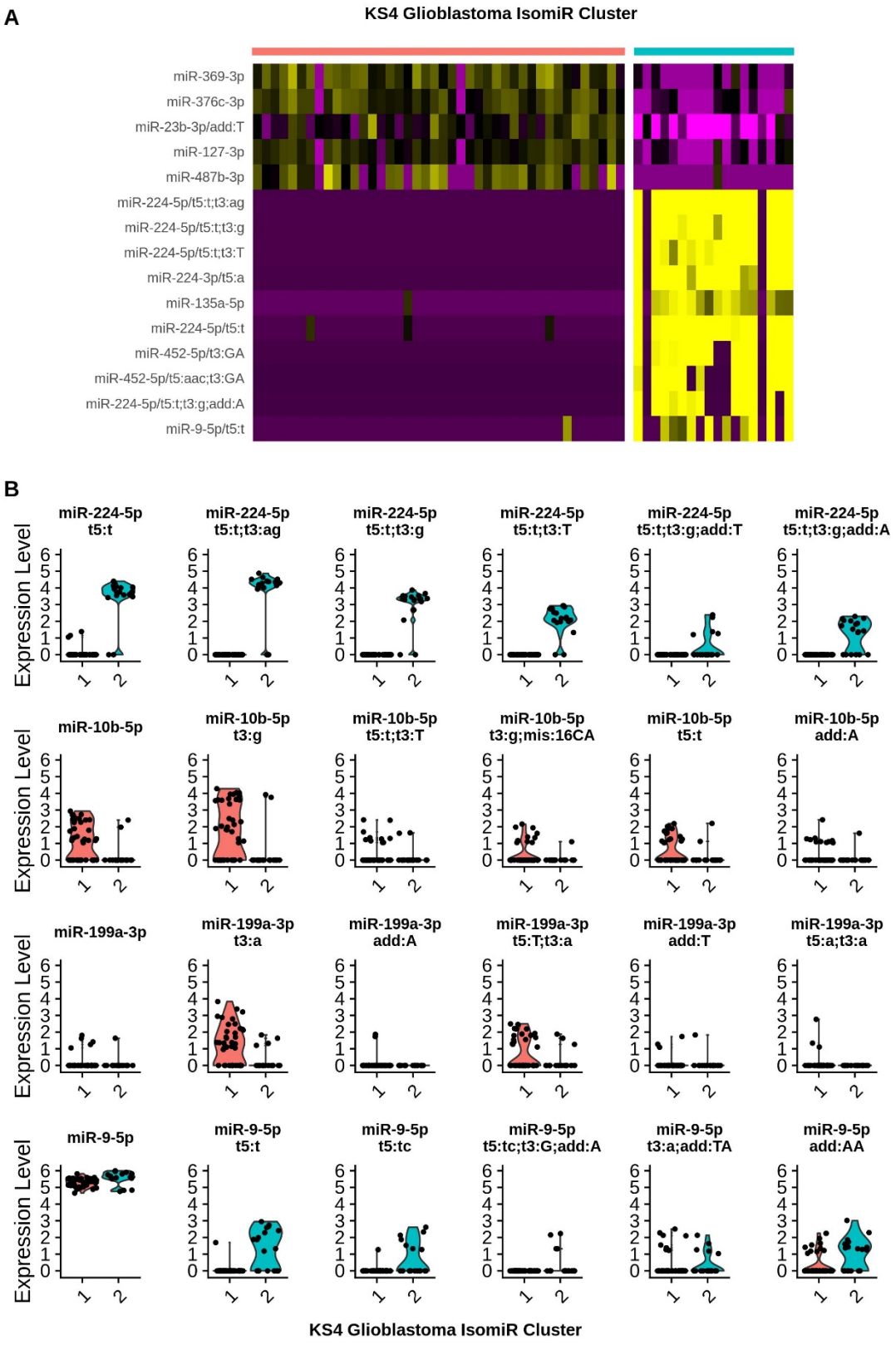

Figure 2.5. Comparison of isomiRs between KS4 Glioblastoma cells in isomiR-based clusters 1 (red) and 2 (blue). A. Heatmap containing isomiRs with the highest fold change expression between clusters. Not all isomiRs shown here are statistically significant. B. Comparison of expression of isomiRs from the same miRNA. isomiR expression shown is in $\log_e$.

2.4. Cell autonomous miRNAs are associated with glioblastoma subtypes and cell state genes

To our knowledge, no paired sequencing data is available that would allow us to directly compare miRNA and RNA expression from single glioblastoma cells. Therefore, to predict the function of the upregulated miRNAs in the two KS4 clusters, we analysed glioblastoma RNA and miRNA expression data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) to identify any association between miRNA expression and marker genes for glioblastoma subtypes or known cell states. The CPTAC data is for bulk sequencing which only measures population level expression and lacks the resolution to characterize cell states in single cells. However, previous work showed that glioblastoma subtypes are indicative of the most dominant cell states within a tumour, with the proneural subtype (TCGA-PN) typically enriched with cells in a neural-progenitor-like (NPC) or oligodendrocyte-progenitor-like (OPC) state, mesenchymal subtype (TCGA-MS) enriched with cells in mesenchymal-like (MES) state, and classical subtype (TCGA-CL) enriched with cells in an astrocyte-like (AC) state[184]. Therefore, we reasoned that if the miRNA pathway had a role in regulating cell states, then it may be possible to identify miRNAs with strong associations to the genes typically upregulated in these cell states using population level data.

2.4.1. Different miRNAs are associated with TCGA glioblastoma subtypes

To detect potential associations between miRNAs and cell states we first scored tumours by their mean expression of marker genes for each glioblastoma subtype[184]. We found that tumours were typically associated with at most one TCGA glioblastoma subtype as they did not score highly in multiple subtypes and there was a negative correlation (Pearson's method) between scores for each subtype (Figure 2.6A-C). Out of the top 200 expressed miRNAs measured across all glioblastoma tumours, 146 miRNAs were significantly correlated with at least one subtype score ($p.adj < 0.05$; Figure 2.8). The strongest correlations of miRNA expression and subtype scores were with the TCGA-PN and TCGA-MS subtypes.

We then compared the correlation coefficients of individual miRNAs across TCGA subtypes and found that miRNAs with the highest positive correlation to a given TCGA subtype were frequently accompanied by a negative correlation to the other subtypes (Figure 2.7 and 2.8). This was most evident when comparing the correlation values of the miRNAs with the TCGA-PN subtype against the TCGA-MS subtype, which revealed a strong anti-correlation pattern (R=-0.69; Figure 2.7B), followed by TCGA-MS against TCGA-CL (R=-0.45) and finally TCGA-PN against TCGA-CL (R=-0.24). Together, this data suggests that miRNA expression is strongly associated with TCGA subtypes and is in line with a potential involvement in regulating cell states.

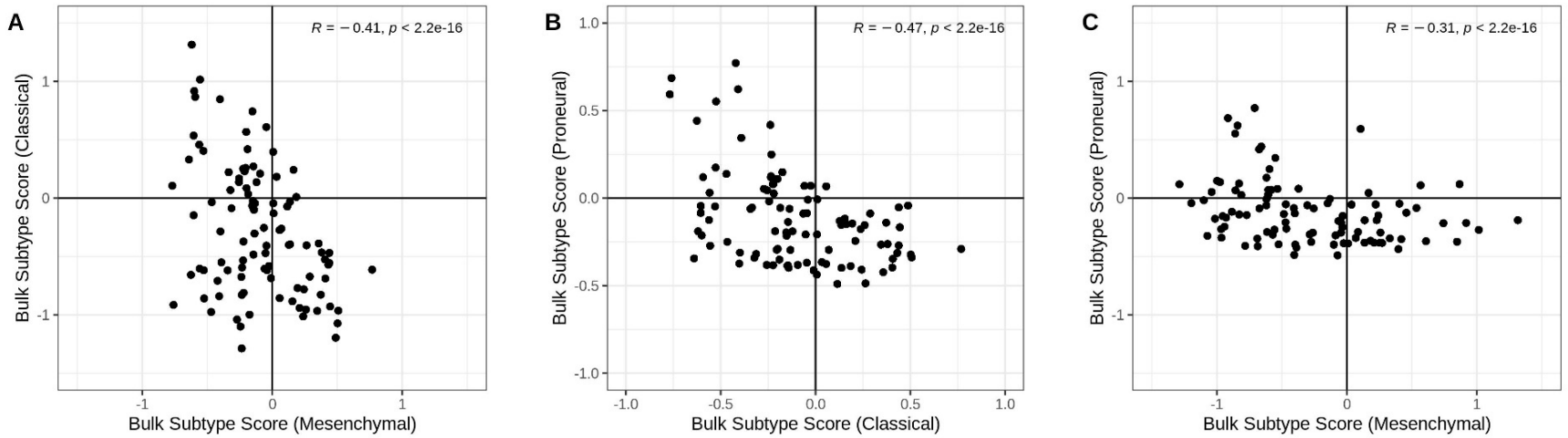

Figure 2.6. Comparison of each pair of TCGA subtype scores in glioblastoma tumours. Scores and Pearson correlation are shown for A. Mesenchymal vs Classical. B. Classical vs Proneural. C. Mesenchymal vs Proneural. Scores were calculated from RNA expression data in CPTAC-3 bulk tumour data, using sets of genes representing each subtype.

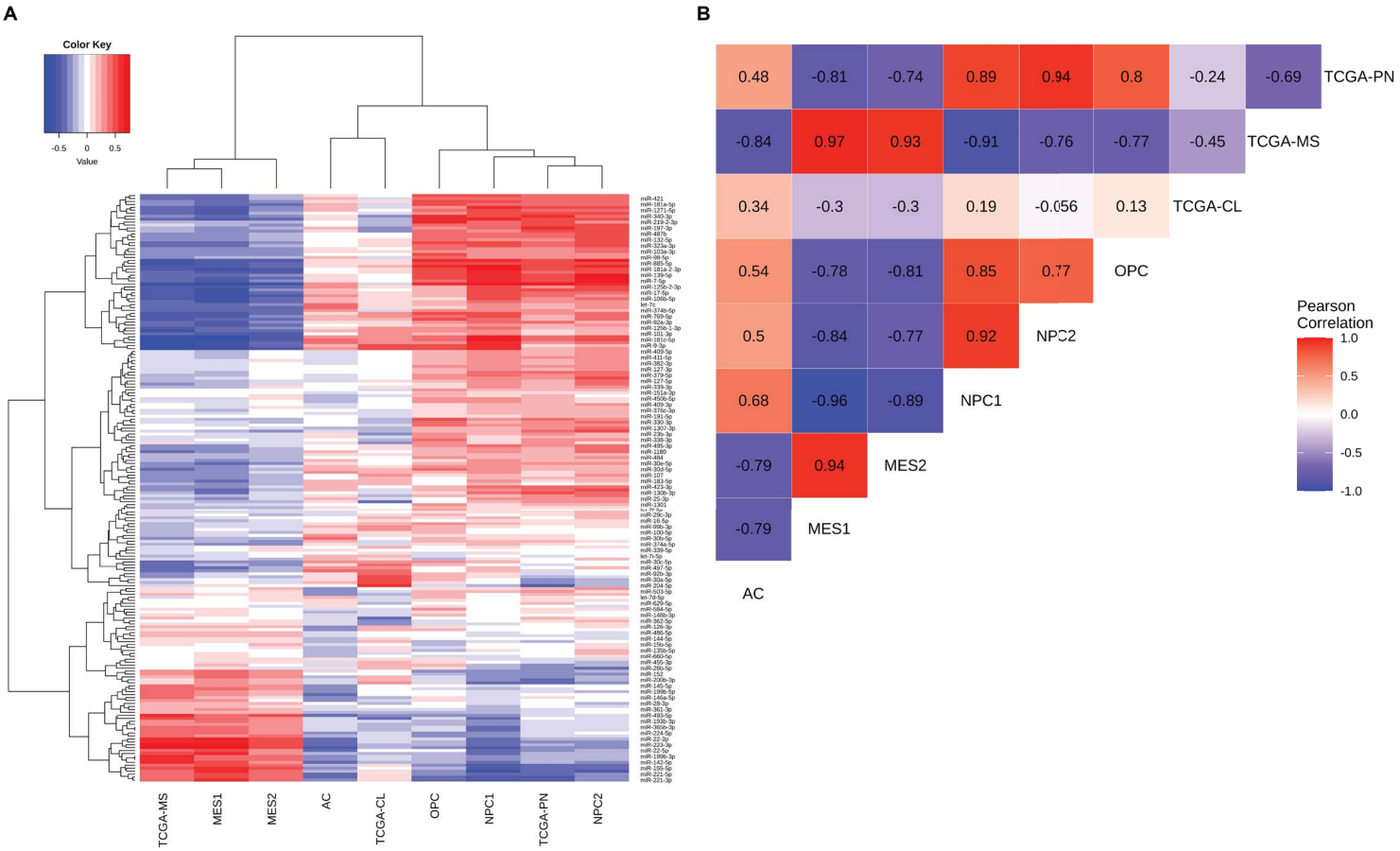

Figure 2.7. Comparison of TCGA subtype and cell state glioblastoma gene modules and their association with miRNAs. **A.** Hierarchical clustering of Pearson correlation values between each miRNA's expression and gene module scores. **B.** Compares miRNA/module correlation values across each pair of gene modules. Gene module scores were calculated from expression of gene sets from Wang et al (for TCGA) and Neftel et al's (for cell states) studies respectively. Two different gene modules were included for MES and NPC states. Pearson's correlation was used to determine correlation metrics. TCGA subtypes – **TCGA-CL:** Classical. **TCGA-MS:** Mesenchymal. **TCGA-PN:** Proneural. Cell states - **AC:** Astrocyte-like. **MES:** Mesenchymal-like. **NPC:** Neural-progenitor-like. **OPC:** Oligodendrocyte-progenitor-like.

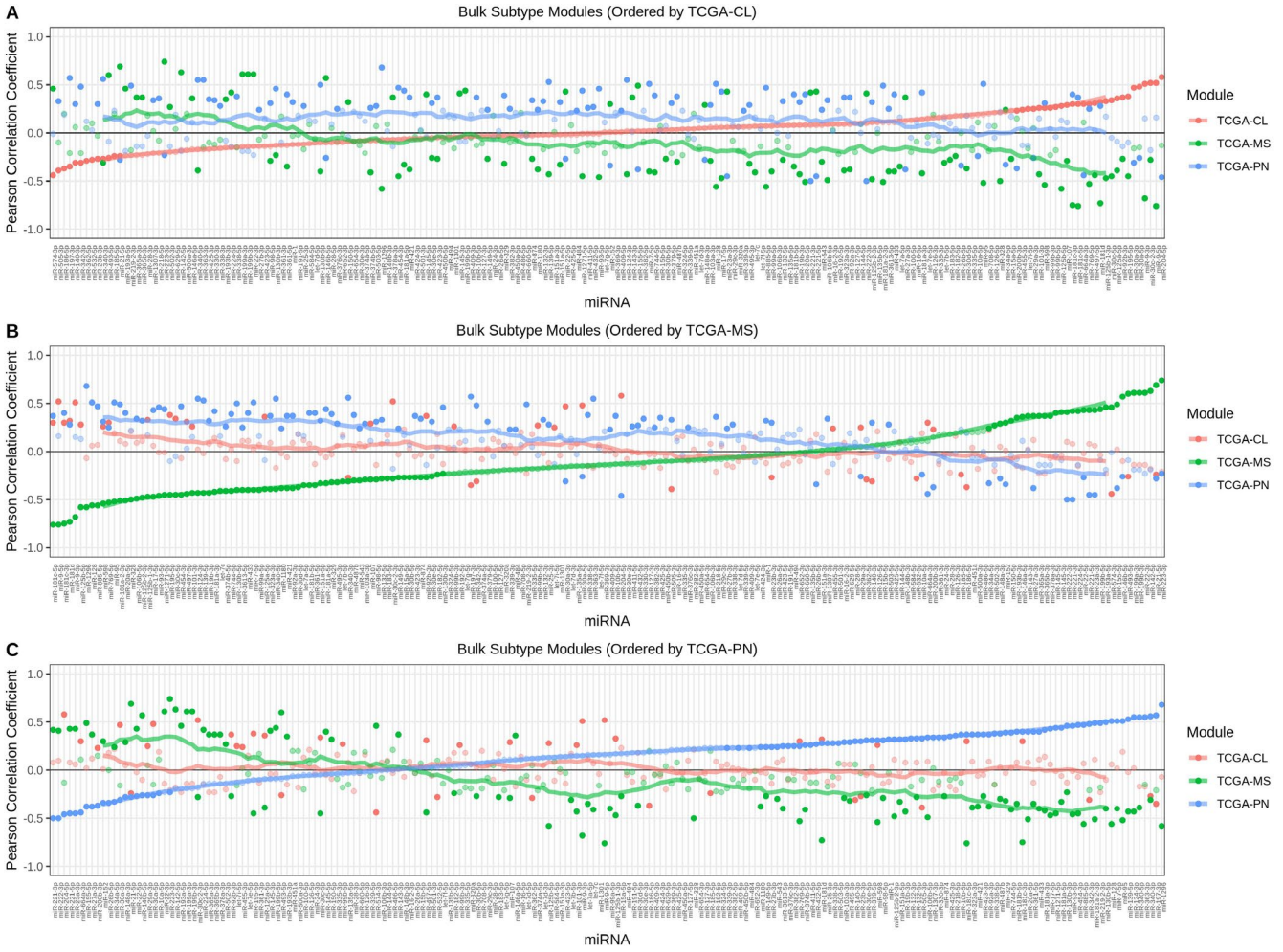

Figure 2.8. Expression of miRNAs correlate with different TCGA subtypes. miRNAs (x-axis) are ordered by ascending Pearson correlation coefficients for the **A.** Classical (TCGA-CL) **B.** Mesenchymal (TCGA-MS) and **C.** Proneural (TCGA-PN) subtypes. Non-significant coefficients are shown as transparent dots. TCGA subtype scores were calculated using gene sets from Wang et al's study[207].

2.4.2. Different miRNAs are associated with glioblastoma cell state genes

Following this we scored tumours by 6 gene modules that represented the 4 cell states described by Neftel et al[184]. This included modules for an astrocyte-like cell state (AC), neural-progenitor-like states (NPC1, NPC2), mesenchymal-like states (MES1, MES2), and an oligodendrocyte-like state (OPC). The highest Pearson correlation values were observed with the NPC1, NPC2, MES1, and MES2 cell states (Figure 2.7A and 2.9). Strong anti-correlation patterns with similar magnitudes were observed between the NPC1/NPC2 states and MES1/MES2 states with most miRNAs. A high number of miRNAs had similar correlations between NPC1, NPC2, OPC, and AC cell states, although were generally weaker with the OPC and AC cell states, and anti-correlation patterns were observed for some miRNAs (Figure 2.7A).

The correlation values were highly similar across the TCGA-Proneural subtype and NPC1/NPC2 states, as well as the TCGA-MS subtype and MES1/MES2 states, as shown with hierarchical clustering (Figure 2.7A) and Pearson correlation (Figure 2.7B). This was consistent with Neftel et al's previous work suggesting these subtypes were predominantly composed of cells exhibiting these respective states[184]. A weaker association was observed between the TCGA-CL subtype and each cell state, with the highest association to the AC state (R=0.34), possibly reflecting this subtypes mixed population of states (Figure 2.7). Collectively, results demonstrate that the TCGA subtypes and cell states, defined by messenger RNA expression, are also evident in miRNA expression and suggest the miRNA pathway may also be involved in their regulation.

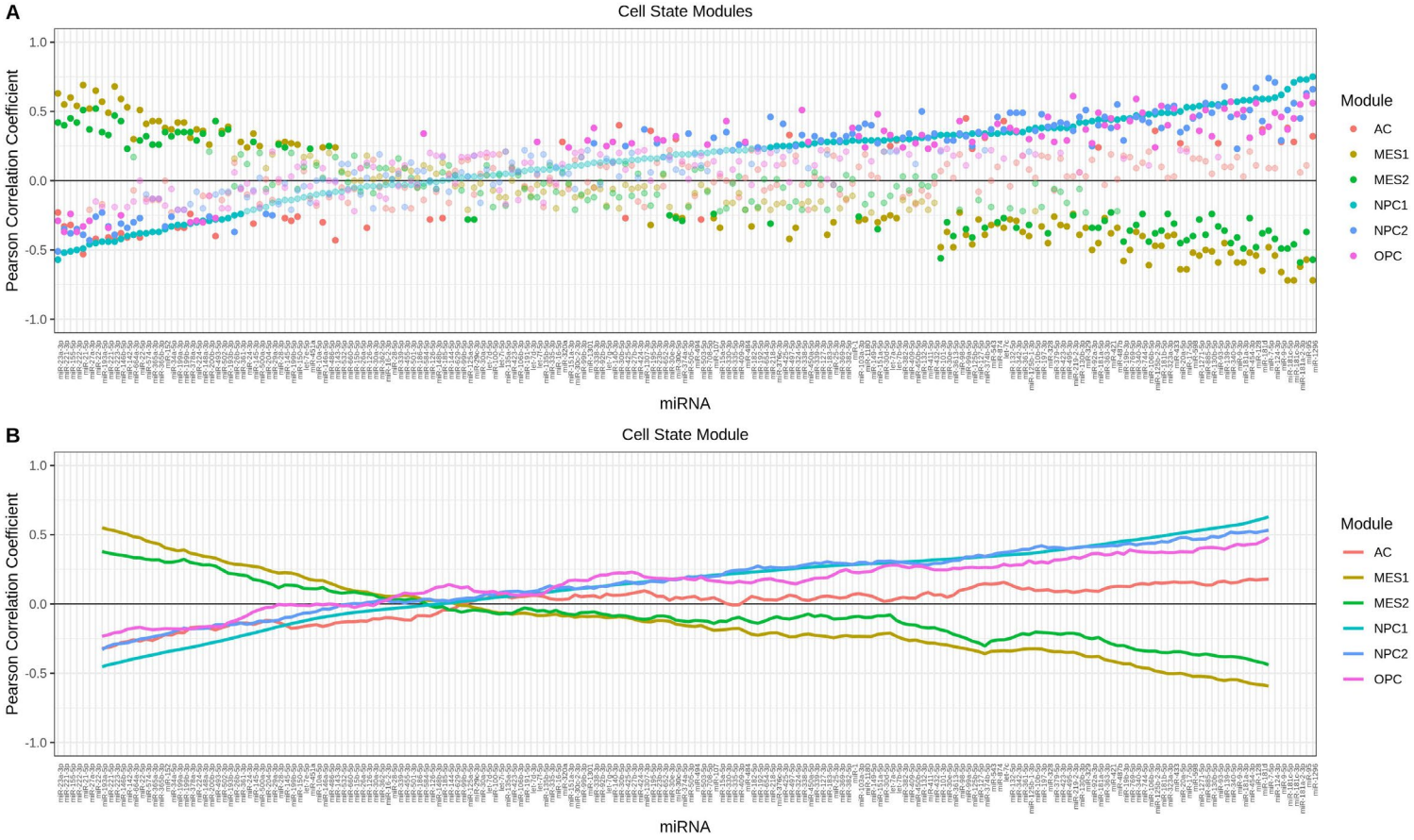

Figure 2.9. Expression of miRNAs correlate with different cell state modules. miRNAs (x-axis) are ordered by ascending Pearson correlation coefficients for the NPC1 cell state. Pearson correlation coefficients are shown for **A.** individual miRNAs or **B.** as rolling averages of the neighbouring 20 miRNAs**.** Cell state expression scores were calculated using gene sets from Neftel et al's study. Two different gene modules were included for MES and NPC states. Non-significant coefficients are shown as transparent dots. **AC:** Astrocyte-like**. MES:** Mesenchymal-like. **NPC:** Neural-progenitor-like. **OPC:** Oligodendrocyte-progenitor-like.

2.4.3. Two miRNA clusters are associated with different cell states in glioblastoma cells

We then focused on the miRNAs that were identified previously as differentially expressed between the two KS4 glioblastoma subpopulations. Most of the miRNAs that were co-upregulated in one of these subpopulations showed a consistency regarding their expression and correlation to the cell state scores (Figure 2.10). For example, 16/20 (p.adj < 0.05) miRNAs upregulated in KS4 cluster 1 (Figure 2.10A) were positively correlated with the NPC1 or NPC2 cell state score and 11/20 (p.adj < 0.05) with the OPC score. Conversely there was a negative correlation with 9/20 (p.adj < 0.05) miRNAs and the MES1 scores. All 18 of the miRNAs from the Dlk1-Dio3 locus had a positive correlation coefficient with the NPC scores however many were weakly correlated or not statistically significant (Figure 2.10A). Only one miRNA, miR-155-5p, appeared to contradict the miRNAs in this group, having a positive correlation with MES1 scores and negative correlations with NPC1, NPC2, and OPC scores. Out of the 5 miRNAs upregulated in KS4 cluster 2 (Figure 2.10B), 3 were positively correlated with the MES1 scores and 2 with the MES2 scores, as well as 2 with the NPC1, and 1 with the NPC2 and OPC scores. All 3 miRNAs from the miR-224/452 cluster negatively correlated with the AC cell state score. The remaining miRNAs did not appear to contradict these observations but lacked statistical significance. We also considered that an aggregated expression of miRNAs from each cluster (Figure 2.10), including all miRNAs belonging to this cluster (All) or only those differentially expressed between the two KS4 glioblastoma subpopulation (DE Only), may provide a stronger association to these cell states than any individual miRNA. We found for both measurements of the Dlk1-Dio3 cluster (Figure 2.10A), their association was similarly positive with the NPC1, NPC2, and OPC states, although not significant with the OPC state for the differentially expressed miRNAs. No significant negative correlation was shown between the aggregated Dlk1-Dio3 miRNA expression and the AC, MES1, and MES2 scores. For the aggregated miR-224/452 cluster miRNA expression (Figure 2.10B), significant positive correlation was observed with the MES1 and MES2 states, and negative correlation with the NPC1, OPC, and AC state scores (both All and DE Only). The evidence indicated that expression of select miRNAs, such as miR-323a-3p and miR-224-5p, may be on par or better than an aggregate expression of the clusters for distinguishing between cell states.

Finally, we noted that the cell state scores were generally consistent with the bulk subtype scores (Figure 2.10), with the miRNAs upregulated in KS4 Cluster 1 generally positively correlating with the TCGA-PN subtype score and miRNAs from KS4 cluster 2 positively correlating with the TCGA-MS subtype score. The combined evidence of cell autonomous regulation of these two miRNA clusters as well as their association with specific cell states highlights a novel form of intra-tumour heterogeneity in glioblastoma and implicates them as potential regulators of glioblastoma cell states.

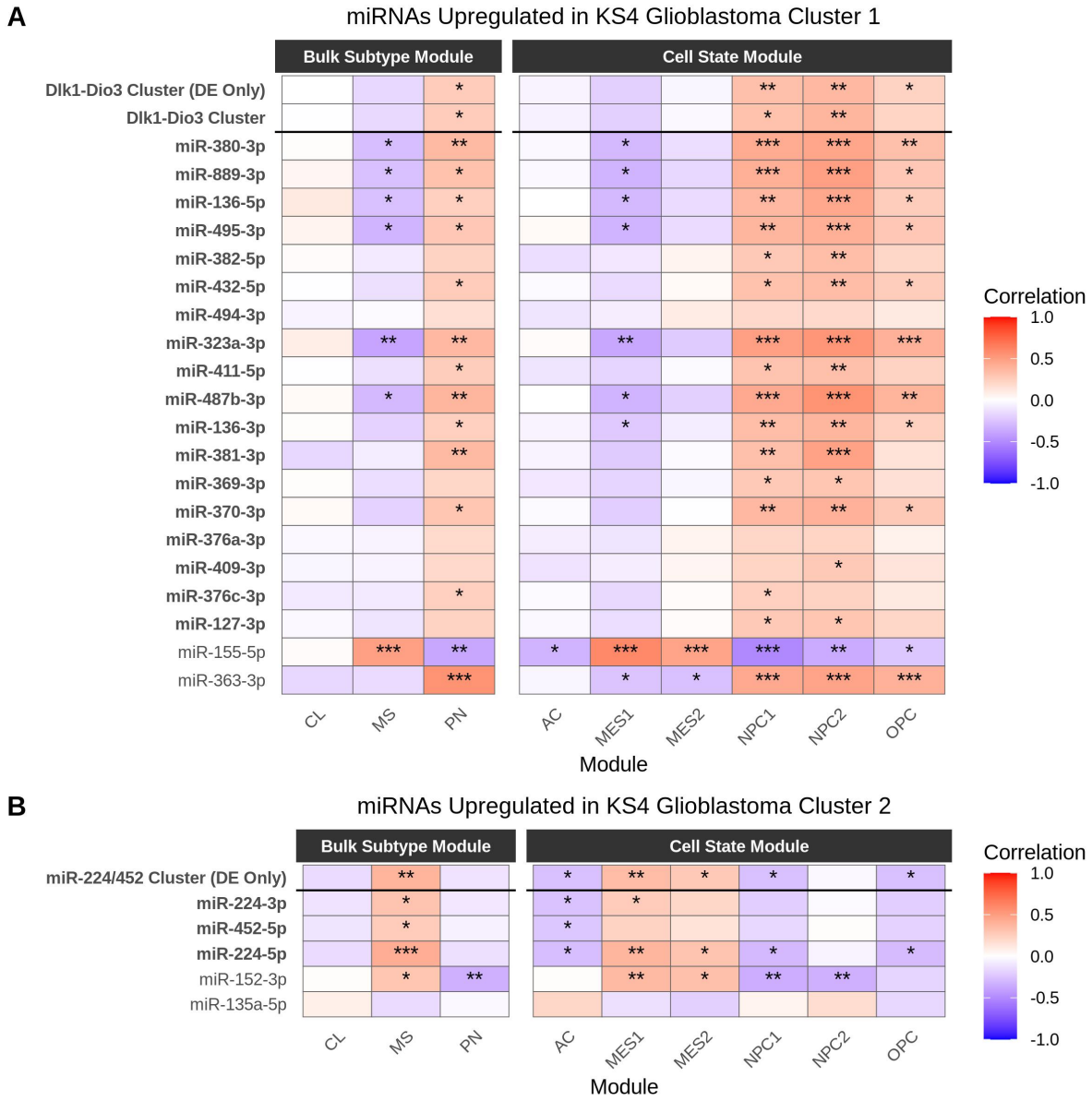

Figure 2.10. Expression of Dlk1-Dio3 and miR-224/452 cluster miRNAs correlate with different cell state modules. **A.** Correlation of the 20 miRNAs upregulated in the KS4 glioblastoma miRNA-based cluster 1 with each TCGA subtype and cell state score. **B.** Correlation of the 5 miRNAs upregulated in the KS4 glioblastoma miRNA-based cluster 2 with each TCGA subtype and cell state score. miRNAs belonging to the Dlk1-Dio3 miRNA cluster **(A)** or miR-224/452 cluster **(B)** are shown in bold. Correlation of combined expression of miRNAs from each cluster is also shown (above black lines), including all miRNAs from cluster (All) or only miRNAs differentially expressed between KS4 cells (DE Only). Pearson's correlation was used to determine correlation metrics. P-values are shown as asterisks - ***:** 0.005 <= p-value < 0.05; ****:** 0.0005 <= p-value < 0.005; *****:** 0.00005 <= p-value < 0.0005.

2.5. The DLk1-Dio3 locus is associated with cell states in single cells

Our previous work identified two miRNA clusters – the Dlk1-Dio3 locus and miR-224/452 cluster – with potential roles in cell state regulation, using population-level data to infer a relationship between these miRNAs and cell state marker genes. Previous studies suggest the miRNAs from the Dlk1-Dio3 locus are co-transcribed from a common primary transcript with 3 long non-coding RNAs – MEG3, MEG8, and MEG9[37]. To expand our study we leveraged single cell RNA-seq data from four separate studies to investigate if there was evidence of an association between the Dlk1-Dio3 locus and glioblastoma cell states on a single cell level[212,213]. Although single cell RNA-seq does not capture miRNAs, we hypothesized that expression of these long non-coding RNAs would be strongly correlated with miRNA expression from this locus and would be potential marker genes for Dlk1-Dio3 miRNA expression in single cells.

2.5.1. Long non-coding RNAs MEG3, MEG8 and MEG9 predict Dlk1-Dio3 miRNA expression in glioblastoma tumours

To determine if the long non-coding RNAs MEG3, MEG8, and MEG9 could predict expression of miRNAs from the Dlk1-Dio3 locus, we used paired RNA and miRNA expression data for glioblastoma tumours from TCGA to determine the correlations of the combined expression of Dlk1-Dio3 miRNAs with each gene in the RNA dataset. There was a high correlation between the combined Dlk1-Dio3 miRNA expression and the non-coding RNAs MEG8 (Pearson's r=0.71), MEG3 (Pearson's r=0.67), and MEG9 (Pearson's r=0.65), as well as with RTL1 (Pearson's r=0.65), all of which are encoded from this locus (Table 2.1). Excluding genes which code for miRNAs or snoRNAs, these four were the most positively correlated genes across the glioblastoma tumours, indicating their expression corresponded well with Dlk1-Dio3 miRNA expression. This provided strong support for their use as markers to infer miRNA expression from this locus when only RNA-seq data is available.

Gene	Pearson's Correlation Coefficient	P-value	Gene in Dlk1-Dio3 locus
MEG8	0.71	1.92E-16	Yes
MEG3	0.67	7.01E-14	Yes
MEG9	0.65	4.88E-13	Yes
RTL1	0.65	5.55E-13	Yes
EIF2S1	0.54	1.18E-08	No
ATP5MJ	0.52	4.07E-08	No
PSMC1	0.50	1.27E-07	No
COA8	0.48	4.20E-07	No
SLC39A9	0.48	6.62E-07	No
UBR7	0.48	5.17E-07	No
BAG5	0.47	1.21E-06	No
CINP	0.47	8.29E-07	No
EIF5	0.47	9.53E-07	No
VTI1B	0.46	1.64E-06	No
AREL1	0.45	3.49E-06	No
FAM131B-AS1	0.45	3.60E-06	No
PPP2R3C	0.45	2.94E-06	No
PPP4R3A	0.45	3.07E-06	No
DRAXIN	0.43	1.29E-05	No
DYNC1H1	0.43	9.72E-06	No

Table 2.1. Genes with the strongest correlation to the combined expression of Dlk1-Dio3 miRNAs, excluding genes encoding for miRNAs (i.e primary or precursor miRNAs) and snoRNAs.

2.5.2. MEG3 expression is associated with the neural progenitor-like and oligodendrocyte progenitor-like cell states in single cells

To investigate if the Dlk1-Dio3 marker genes were associated with any cell states we utilized single cell RNA-seq data from 4 different studies, which included 43 glioblastoma tumours and 16269 cells after filtering[184,212,213,226]. For the Neftel study we analysed the tumours as two distinct datasets depending on whether they were sequenced by the Smart-seq2 (labelled Neftel dataset) or 10X Genomics (labelled Neftel 10X dataset) methods, and some of the tumours featured in multiple datasets. Of the Dlk1-Dio3 markers, only MEG3 was detectable in a high number of cells so we focused on this gene.

Each glioblastoma cell was assigned a score for each cell state, using the same methodology described previously. We then calculated the Pearson correlation between the cell state module scores and MEG3 expression (log_2[TPM+1]) for all cells in each dataset. There was a significant positive correlation between MEG3 expression and the NPC1 and NPC2 cell state scores found in all datasets (Figure 2.11). Correlations ranged from positively weak to moderate for the NPC1 (Pearson's r=0.07 to 0.53) and NPC2 scores (Pearson r=0.08 to 0.61). Similar observations, albeit to a lesser extent, were seen with the OPC scores (Pearson's r=0.00 to 0.48). Additionally, there was a significant weak to moderately negative correlation between MEG3 expression and the MES1 scores (Pearson's r=-0.12 to -0.42) across datasets.

To determine if this observation was consistent across individual tumours, we separated glioblastoma cells accordingly (Figure 2.12). We found that for most tumours (38/50), MEG3 expression was positively correlated with at least one of the NPC cell state scores, and many were also positively correlated with the OPC scores (22/50). Most tumours with a positive correlation of MEG3 expression and NPC scores also had a corresponding negative correlation with the MES1 and/or MES2 scores, again observable across multiple datasets. The coefficients were generally stronger in tumours with a higher expression of MEG3, suggesting that tumours with lower expression of MEG3 either had insufficient statistical power to detect this association or this association is not present in all glioblastomas. Although we did not observe any strong correlation (Pearson's R >= 0.7) with any modules, the observations were consistent with the Dlk1-Dio3 miRNAs from the TCGA data.

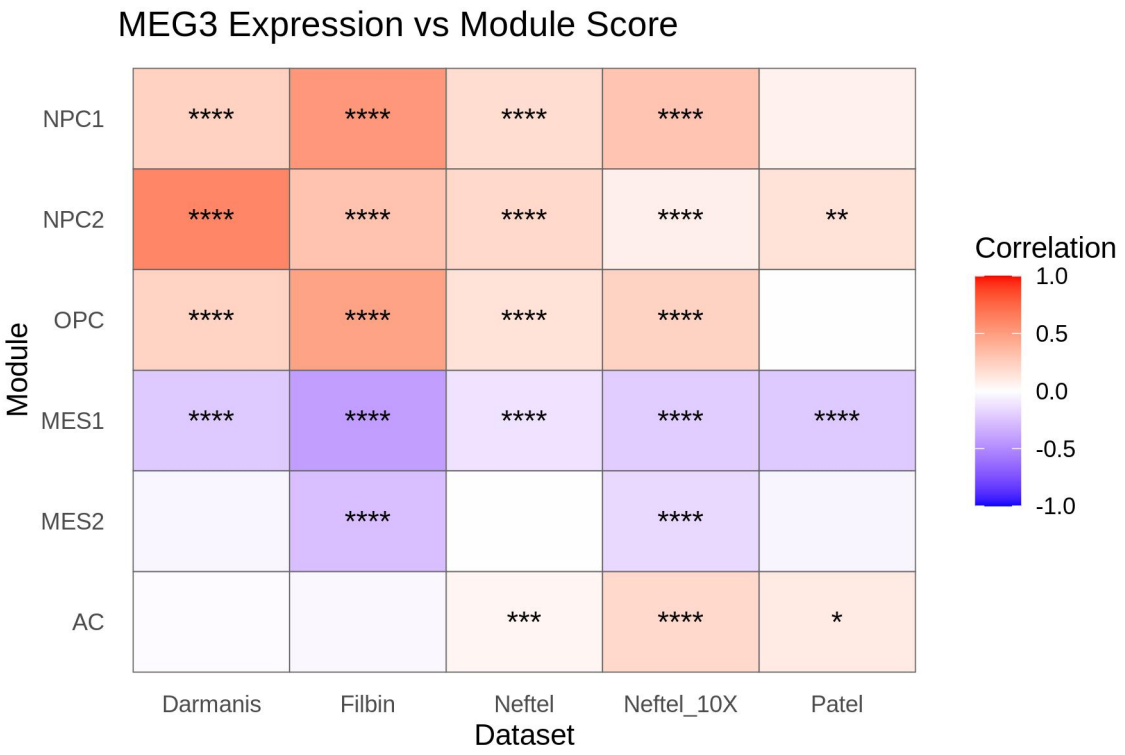

Figure 2.11. Correlation of MEG3 expression with cell state module scores across all glioblastoma cells for each dataset. Scores were calculated from expression of gene modules highly expressed in their respective glioblastoma cell state. Scoring method and gene modules were obtained from the Neftel et al study[184]. Pearson's method was used for correlation, and p-value significance is shown as asterisks. ***:** 0.005 <= p-value < 0.05; ****:** 0.0005 <= p-value < 0.005; *****:** 0.00005 <= p-value < 0.0005; ******:** 0.000005 <= p-value < 0.00005.

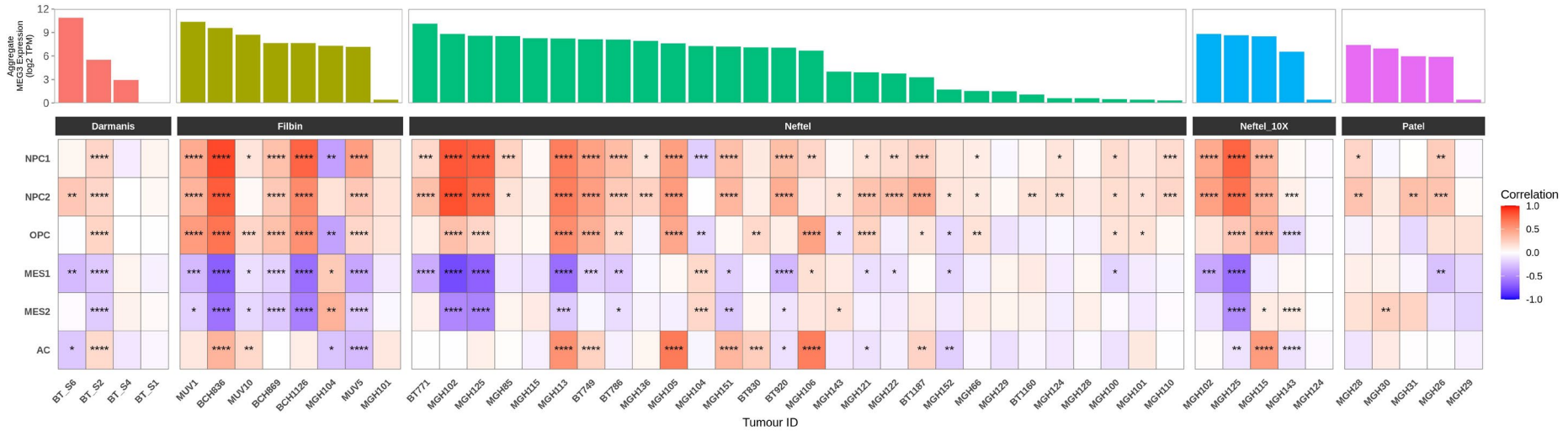

Figure 2.12. MEG3 expression and correlation with cell state module scores in cells from each glioblastoma tumour. Scores were calculated from expression of gene modules highly expressed in their respective glioblastoma cell state. Scoring method and gene modules were obtained from the Neftel et al study[184]. Pearson's method was used for correlation, and p-value significance is shown as asterisks. ***:** 0.005 <= p-value < 0.05; ****:** 0.0005 <= p-value < 0.005; *****:** 0.00005 <= p-value < 0.0005; ******:** 0.000005 <= p-value < 0.00005.

2.6. Target prediction identifies Dlk1-Dio3 miRNAs regulate key transcription factors and regulatory pathways involved in cancer

As miRNAs function by regulating gene targets, we used miRNAtap to predict the targets of miRNAs from the Dlk1-Dio3 and miR-224/452 clusters and compared them to the cell state gene modules to see if any cell states were highly represented (Figure 2.13). We found that the proportion of cell state genes that were targets of at least one the 18 upregulated miRNAs from the Dlk1-Dio3 locus were not substantially different, ranging from 18-28% (Figure 2.13A). The cell state modules with the highest proportion of targets were the NPC1 and NPC2 modules, where 28.0% (14/50) of genes were identified as Dlk1-Dio3 targets. This was followed by 25.6% (10/39) in the AC module, 24.0% (12/50) in the MES2 module, 22.0% (11/50) in the MES1 module, and finally 18.0% (9/50) in the OPC module. As the number of targets across all 18 miRNAs (n=3605) represented a significant portion of overall genes, we also considered genes which were targets of multiple miRNAs from the same cluster, which we expected would be more strongly regulated (Figure 2.13). When only considering genes targeted by 2 or more miRNAs from Dlk1-Dio3 locus, the number of targets matching cell state genes dropped for all cell state modules, however the NPC1 module remained the most highly represented (6/50 genes). When increasing to 3 or more miRNAs, only 1 gene remained from the NPC2 and OPC modules, and no other cell states had matching targets. Interestingly, the NPC2 marker gene NFIB was a predicted target of 7 of the 18 identified Dlk1-Dio3 miRNAs, suggesting they downregulated this gene. This was unexpected given that our previous work indicated the miRNAs from this cluster were upregulating the genes which represented this state. Nonetheless, the results highlighted that regulation of this gene is likely an important function of this miRNA cluster.

We also investigated targets of the 3 miRNAs from the miR-224/452 cluster (Figure 2.13B). The proportion of cell state genes that were targets of at least one miRNA from this locus were similar across most cell states. Representation was highest in the NPC2 module (7/50), followed by the NPC1 module (5/50), then MES1, MES2 and OPC modules (4/50), and finally the AC module (3/39). Only the NPC2 cell state module had a gene, MAP1B, that was targeted by 2 or more miRNAs from this cluster.

Following this we considered whether the function of either miRNA cluster could be predicted through KEGG pathway enrichment analysis of their gene targets (Figure 2.14). For the Dlk1-Dio3 miRNA targets there was a significant enrichment in pathways in cancer (ID: hsa05200) as well as several pathways associated with cell signalling, including Proteoglycans in cancer (ID: hsa05205), Hippo signalling pathway (ID: hsa04390), FoxO signalling pathway (ID: hsa04068), and MAPK signalling pathway (ID: hsa04010), which have roles in cell adhesion, proliferation, differentiation, and apoptosis (Figure 2.14A)[234–237]. miRNA targets from the miR-224/452 cluster were also significantly enriched in

multiple signalling pathways which included the Wnt signalling pathway (ID: hsa04310) and Hedgehog signalling pathway (ID: hsa04340), important regulators of stem cell renewal and epithelial-mesenchymal transition (Figure 2.14B)[238,239]. Although many of the signalling pathways mentioned above may be involved in cell state regulation, we were unable to find strong evidence that miRNA targets from the Dlk1-Dio3 locus or miR-224/452 cluster had a specific role in any known cell states.

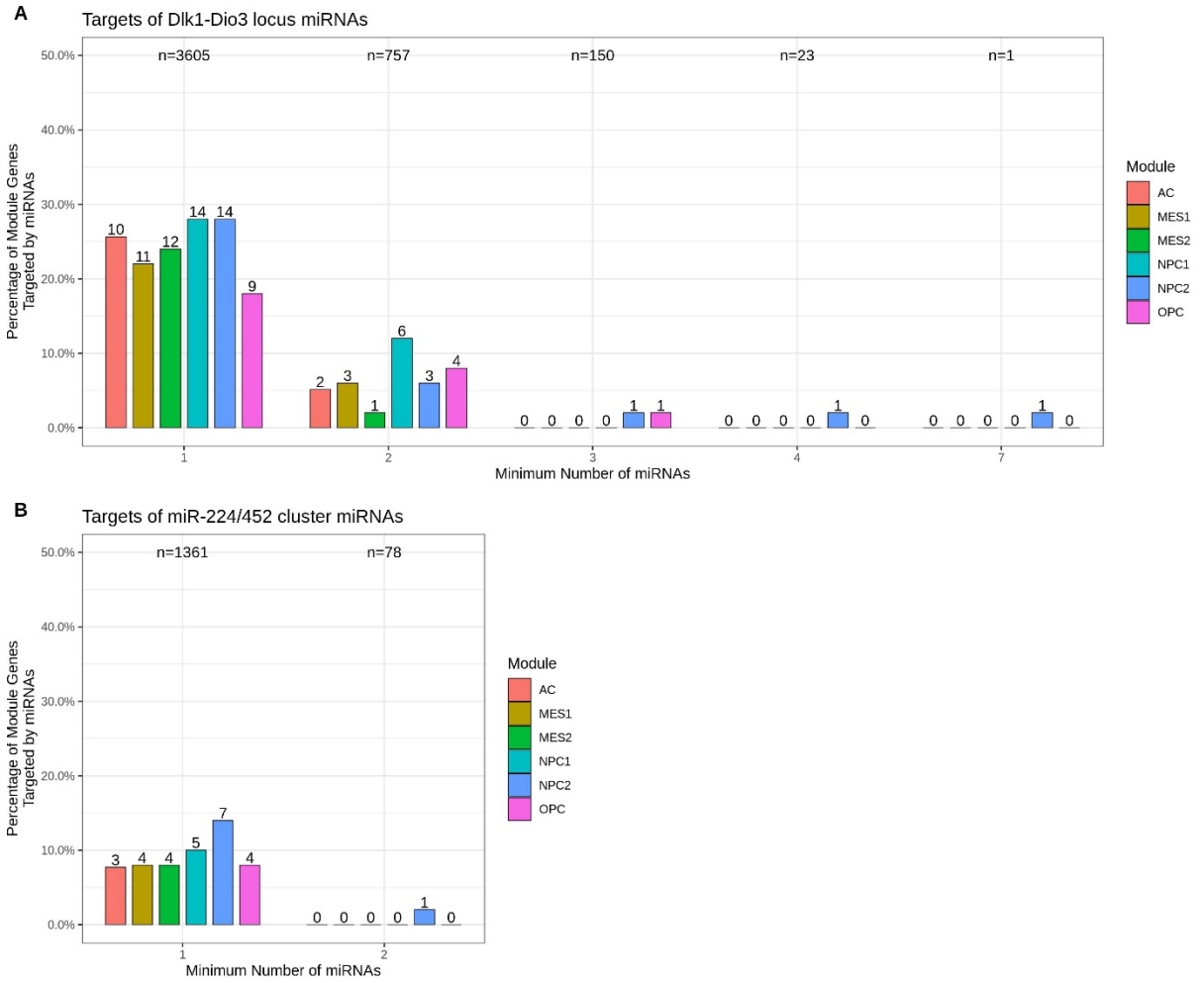

Figure 2.13. Common genes between cell state modules and miRNA targets. Shows predicted targets for **A.** Dlk1-Dio3 locus miRNAs and **B.** miR-224/452 cluster miRNAs. We considered all predicted targets from cluster miRNAs, as well as only those targeted by multiple miRNAs (x-axis). Only includes miRNAs from each cluster which were differentially expressed in the KS4 cells. **AC:** Astrocyte-like. **MES:** Mesenchymal-like. **NPC:** Neural-progenitor-like. **OPC:** Oligodendrocyte-progenitor-like.

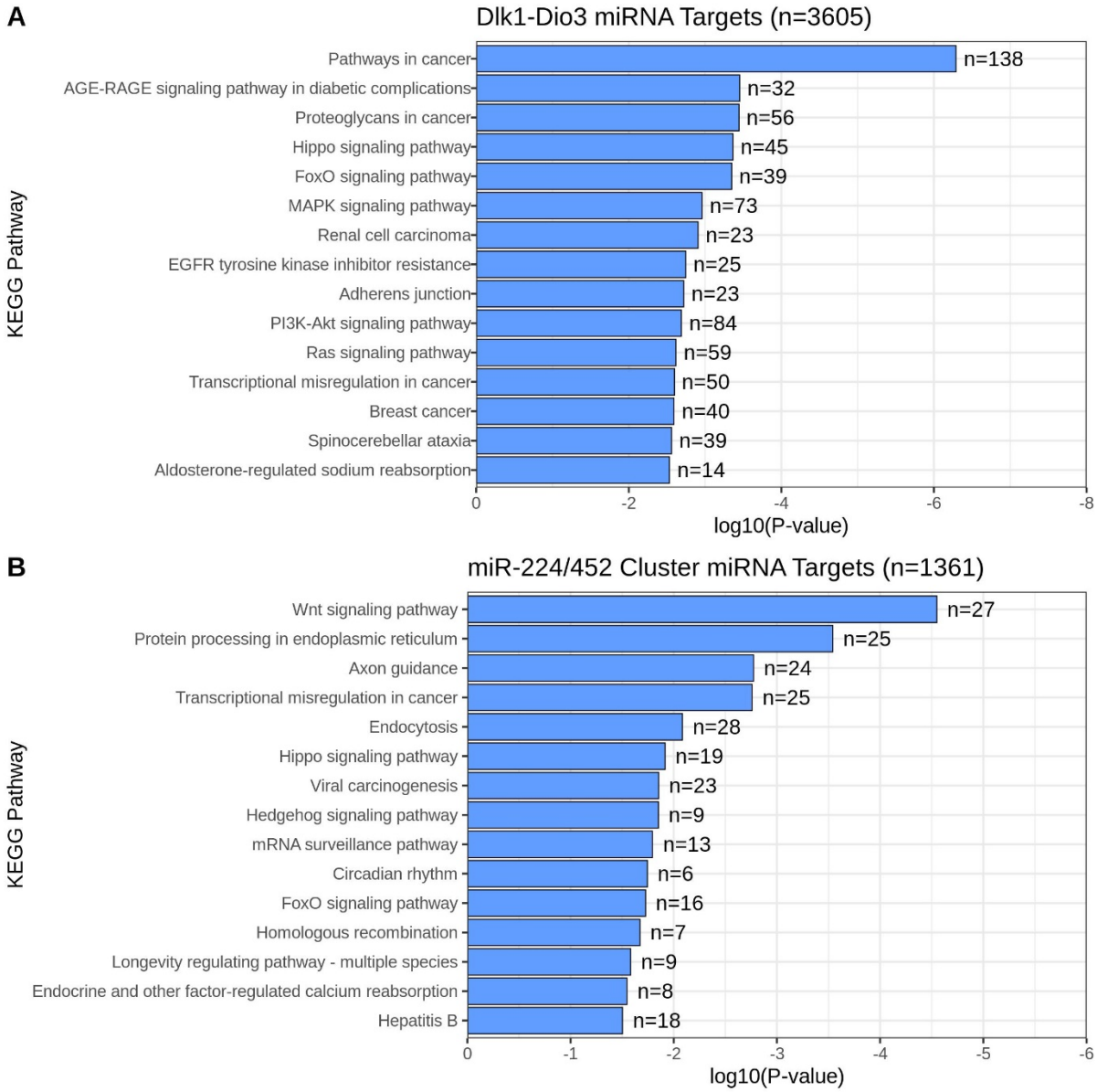

Figure 2.14. Gene Set Enrichment Analysis of KEGG pathways for miRNA cluster targets. **A.** Targets of the Dlk1-Dio3 miRNAs. **B.** Targets of the miR-224/452 miRNAs. Only includes miRNAs from each cluster which were differentially expressed in the KS4 cells.

2.6.2. Expression of miRNA targets is associated positively with NPC and OPC cell states and negatively with MES states

One of the primary mechanisms of gene regulation in the miRNA pathway is to facilitate degradation of transcriptional targets. Therefore, we hypothesized if miRNAs were important in the regulation of any cell states, their activity would be detectable through a change in expression of their targets. For example, cells with higher expression of miRNAs should have, on average, lower expression of their targets and conversely cells with lower expression of miRNAs should have, on average, higher expression of their targets. To investigate the regulatory effects of miRNAs from the Dlk1-Dio3 locus or miR-224/452 clusters with cells in each of the cell states, we calculated the mean expression of targets within each cell and then binned cells as high (score > 1), none (-1 > score >= 1), or low (score <= -1) scoring for each cell state score calculated in the previous section (Figure 2.15 and 2.16). As several factors, including library preparation and sequencing method are known to influence gene quantification, we treated cells from different datasets separately.

With the Dlk1-Dio3 miRNA targets, there was typically a high amount of variation in target expression across cells even within the same scoring bin (Figure 2.15). We observed a minor but statistically significant increase in target expression when comparing low to high scoring MES1 (in 5/5 datasets) and MES2 (in 4/5 datasets) cells. For the other cell state modules, there was a general lack of consistency in results when comparing cells from different datasets, indicating no association between target expression and cell state score. Similar results were observed with the miR-224/452 cluster targets, which also had a high variation in expression with cells in each scoring bin. For both the MES1 and MES2 cell states, there was a similar trend with increased expression in the low scoring cells in the Darmanis, Neftel, and Filbin datasets and no significant difference detected in the remaining datasets. Additionally, no clear trend was evident when examining the other cell states for the miR-224/452 cluster targets (Figure 2.16).

We considered that the high stochastic variability of gene expression between cells may confound any measurable changes in target expression between high and low scoring cell states. The scoring method used previously for the cell states effectively measures mean expression of associated marker genes, while also adjusting values by expression of a set of control genes of each cell. Therefore, to account for stochastic variability we used this method to score cells according to their expression of gene targets from the Dlk1-Dio3 or miR-224/452 cluster miRNAs and refer to them as cluster target scores. The difference between cells with high and low cell state scores was clearer when comparing target scores instead of expression (Figure 2.17 and 2.18). The Dlk1-Dio3 (Figure 2.17) and miR-224/452 (Figure 2.18) target scores were both increased in the high scoring MES1 and MES2 cells, and decreased in high scoring NPC1, NPC2. A decrease in miRNA target scores was also observed in high

scoring OPC and AC cells however this was less pronounced in the miR-224/452 target scores. Together this data suggested there was a measurable change in the expression of targets by the Dlk1-Dio3 and miR-224/452 miRNA cluster when comparing cells in different cell states, and this may indicate that the Dlk1-Dio3 miRNA cluster is more active in MES cells and less active in NPC cells.

Finally, we calculated the Pearson correlation of target scores and cell state scores, calculating coefficients for each tumour separately (Figure 2.19). For the Dlk1-Dio3 cluster targets, there was a weak positive correlation between the AC, NPC1, NPC2, and OPC scores and a weak negative correlation with MES1 scores for cells in most of the glioblastoma tumours (Figure 2.19A). We also scored cells using genes targeted by multiple miRNAs for this cluster (up to a maximum of 7) and observed a high degree of similarity when including all targets (Figure 2.19A). A median weak positive correlation for the AC, NPC1, NPC2, and OPC scores and median weak negative correlation for the MES1 scores was also observed with targets from the miR-224/452 cluster (Figure 2.19B). These results suggested that downregulation of the Dlk1-Dio3 cluster targets was not evident in cells scoring highly for the NPC1, NPC2, OPC and AC modules, despite our previous work indicating an upregulation of this cluster's miRNAs. For the miR-224/452 cluster targets, the observed negative correlation with the MES1 module (genes targeted by at least one miRNA from this cluster) and MES2 (only with genes targeted by at least 2 miRNAs from this cluster) was consistent with a regulatory effect from this cluster's miRNAs on the mesenchymal cell state.

To investigate if this trend was exclusive to these miRNA clusters, we expanded these results to include targets from each of the top 200 expressed miRNAs across TCGA glioblastoma tumours (Figure 2.20). We found that there was a general trend for targets of miRNAs to be positively correlated with the NPC1, NPC2, and OPC scores as well as negatively correlated with the MES1 and MES2 scores, indicating the results were not specific to these two miRNA clusters. The correlation coefficients for the AC scores were more mixed. Notably, target scores for miRNAs with a stronger positive correlation with the NPC1, NPC2, and OPC scores nearly always corresponded with a stronger negative correlation to the MES1 and MES2 scores, mirroring previous observations of an inverse relationship between NPC/OPC and MES states when comparing miRNA expression and cell state scores (Figure 2.9).

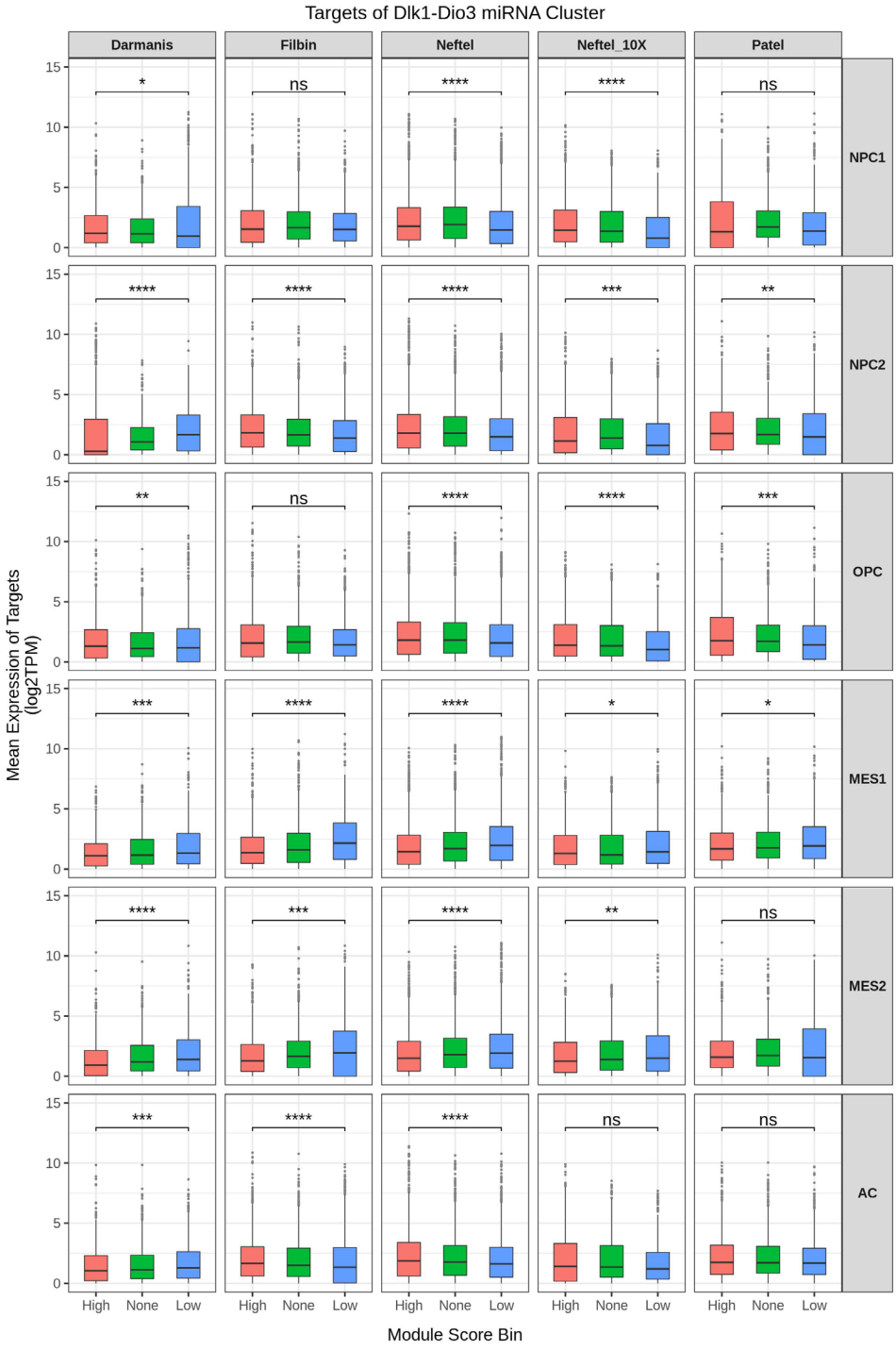

Figure 2.15. Expression of gene targets for miRNAs from the Dlk1-Dio3 cluster. For each cell state score, glioblastoma cells were divided into high (score > 1), none (-1 > score >= 1), or low (score <= -1) scoring bins and mean expression of targets was compared between high and low bins using the Student's t-test. Data is shown from 5 different datasets – Darmanis, Filbin, Neftel, Neftel_10X, and Patel. Significance shown as asterisks - ***:** 0.005 <= p-value < 0.05; ****:** 0.0005 <= p-value < 0.005; *****:** 0.00005 <= p-value < 0.0005; ******:** 0.000005 <= p-value < 0.00005. **AC:** Astrocyte-like. **MES:** Mesenchymal-like. **NPC:** Neural-progenitor-like. **OPC:** Oligodendrocyte-progenitor-like.

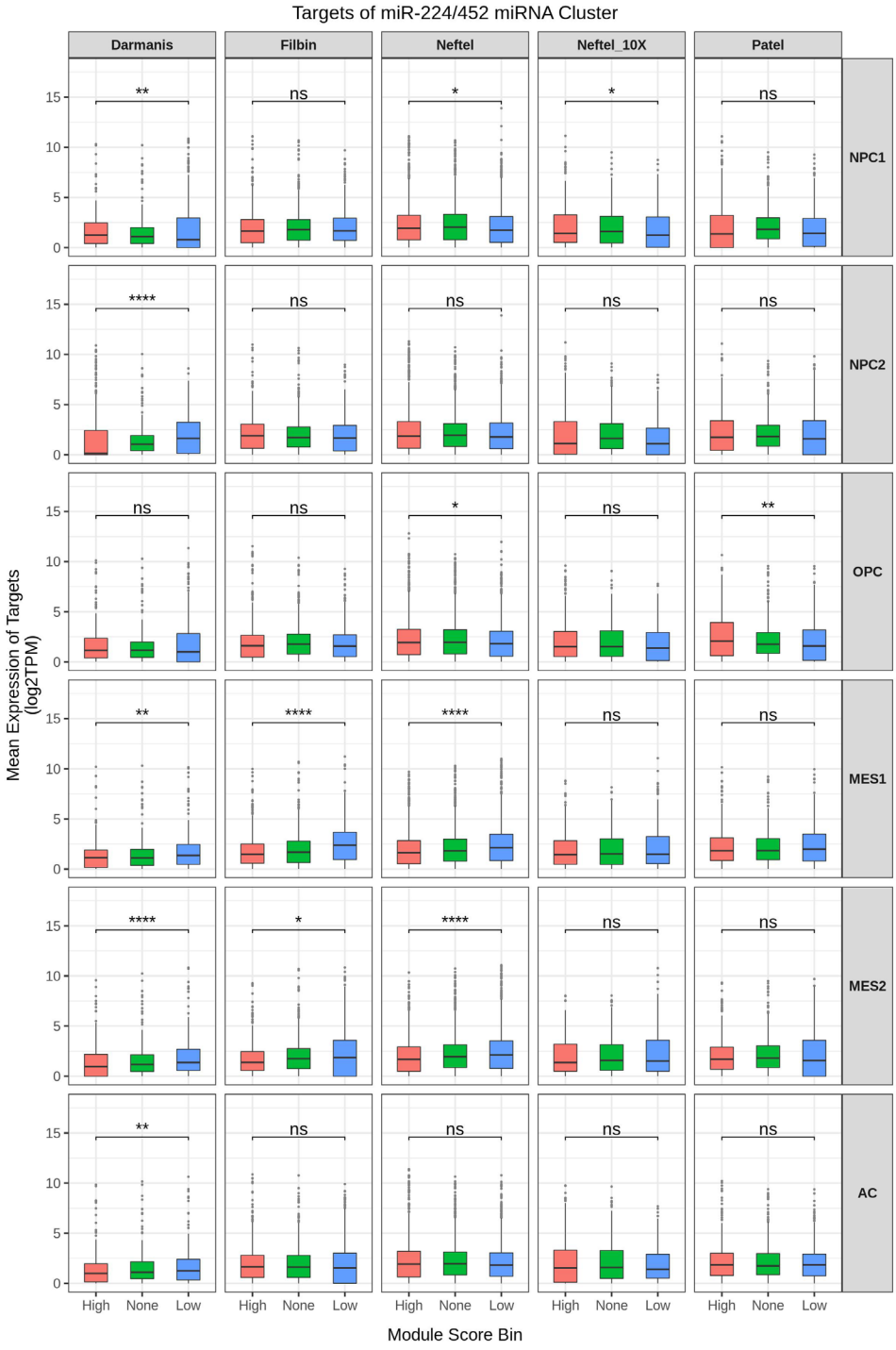

Figure 2.16. Expression of gene targets for miRNAs from the miR-224/452 cluster. For each cell state score, glioblastoma cells were divided into high (score > 1), none (-1 > score >= 1), or low (score <= -1) scoring bins and mean expression of targets was compared between high and low bins using the Student's t-test. Data is shown from 5 different datasets – Darmanis, Filbin, Neftel, Neftel_10X, and Patel. Significance shown as asterisks - ***:** 0.005 <= p-value < 0.05; ****:** 0.0005 <= p-value < 0.005; *****:** 0.00005 <= p-value < 0.0005; ******:** 0.000005 <= p-value < 0.00005. **AC:** Astrocyte-like. **MES:** Mesenchymal-like. **NPC:** Neural-progenitor-like. **OPC:** Oligodendrocyte-progenitor-like.

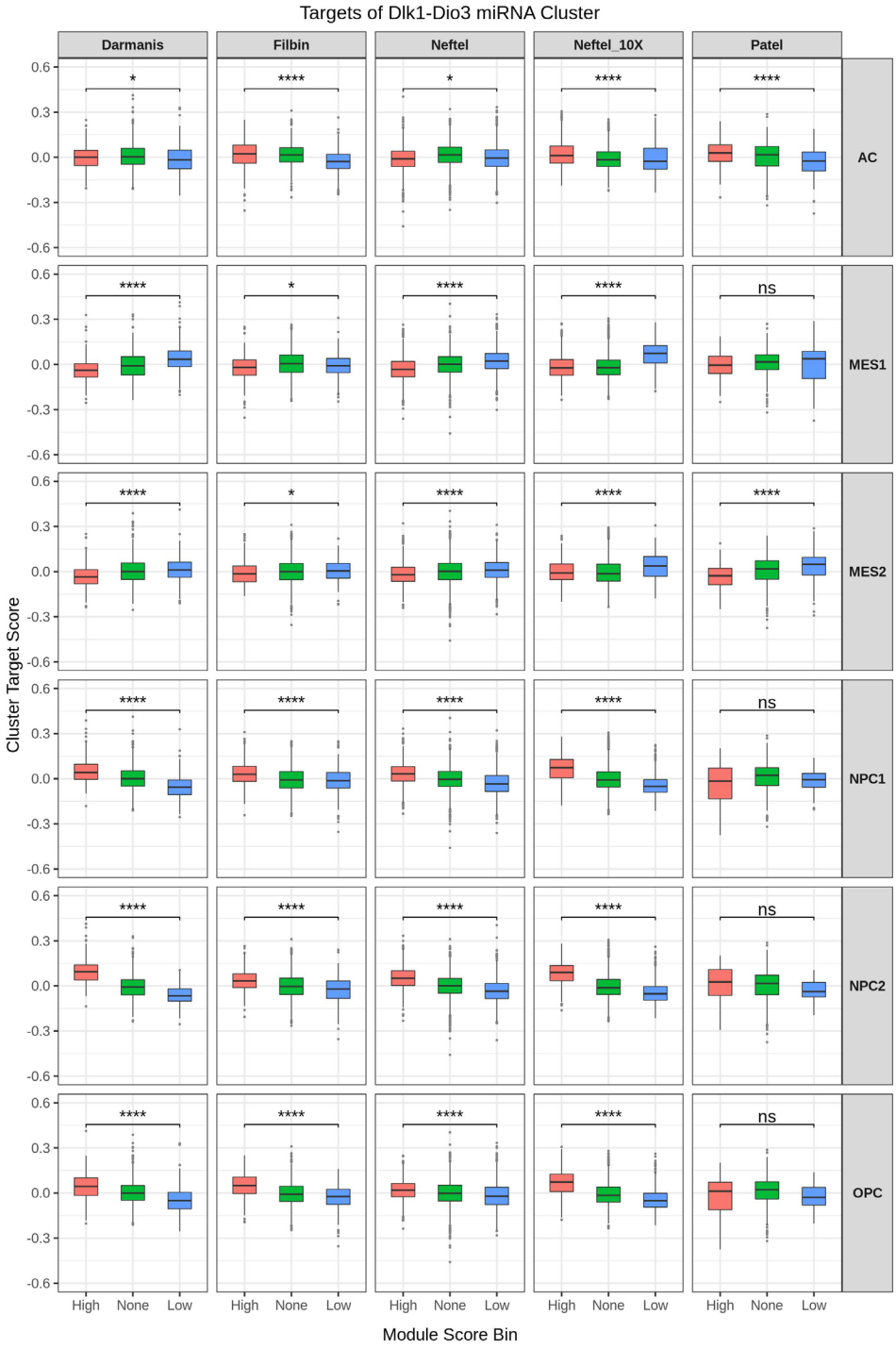

Figure 2.17. Cluster target scores for the Dlk1-Dio3 cluster. Target scores were calculated using the same scoring method for cell states. For each cell state score, glioblastoma cells were divided into high (score > 1), none (-1 > score >= 1), or low (score <= -1) scoring bins and mean expression of targets was compared between high and low bins using the Student's t-test. Data is shown from 5 different datasets – Darmanis, Filbin, Neftel, Neftel_10X, and Patel. Significance shown as asterisks - *****: 0.005 <= p-value < 0.05; ******: 0.0005 <= p-value < 0.005; *******: 0.00005 <= p-value < 0.0005; ********: 0.000005 <= p-value < 0.00005. **AC:** Astrocyte-like. **MES:** Mesenchymal-like. **NPC:** Neural-progenitor-like. **OPC:** Oligodendrocyte-progenitor-like.

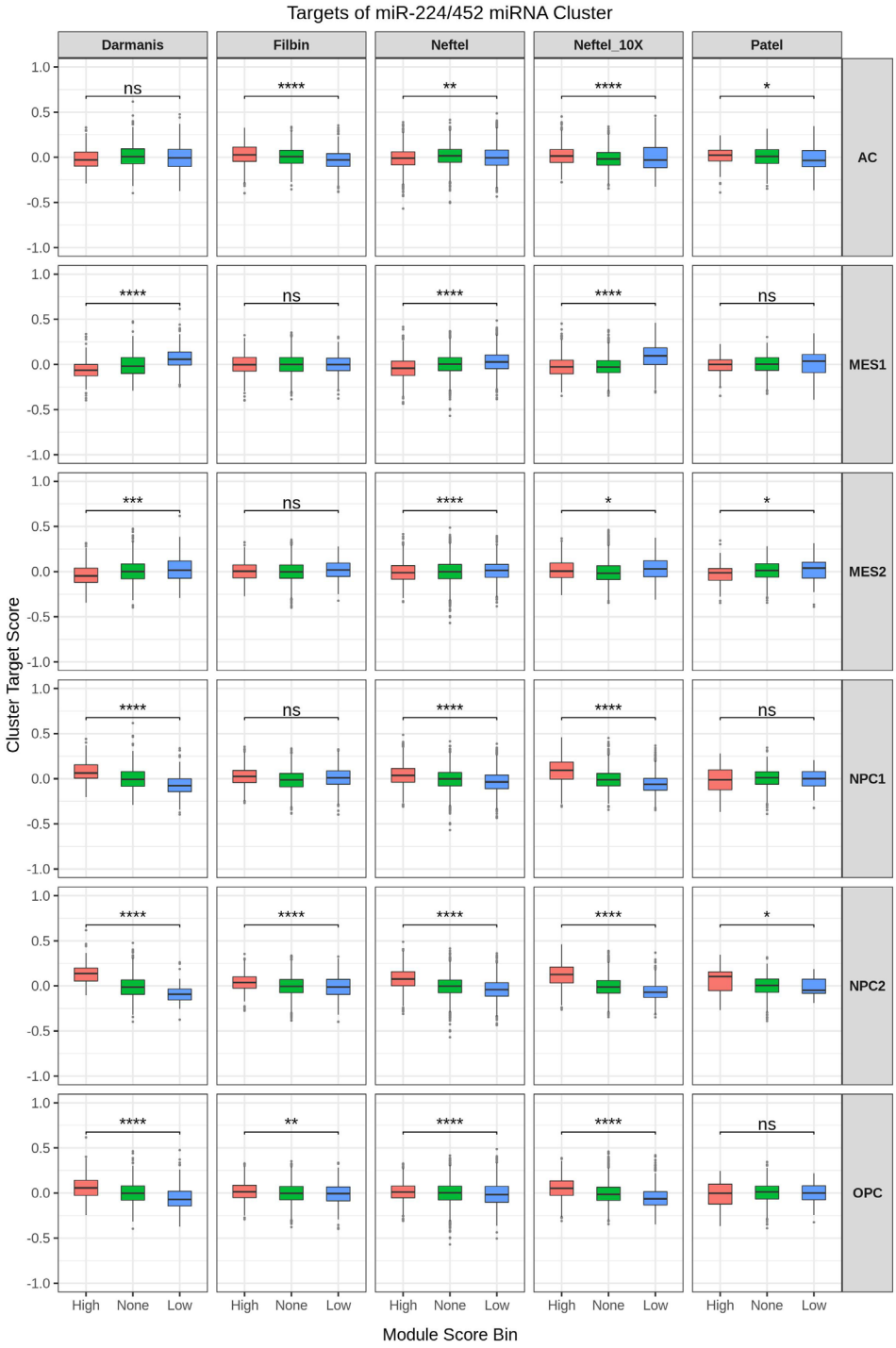

Figure 2.18. Cluster target scores for the miR-224/452 cluster. Target scores were calculated using the same scoring method for cell states. For each cell state score, glioblastoma cells were divided into high (score > 1), none (-1 > score >= 1), or low (score <= -1) scoring bins and mean expression of targets was compared between high and low bins using the Student's t-test. Data is shown from 5 different datasets – Darmanis, Filbin, Neftel, Neftel_10X, and Patel. Significance shown as asterisks - *****: 0.005 <= p-value < 0.05; ******: 0.0005 <= p-value < 0.005; *******: 0.00005 <= p-value < 0.0005; ********: 0.000005 <= p-value < 0.00005. **AC:** Astrocyte-like. **MES:** Mesenchymal-like. **NPC:** Neural-progenitor-like. **OPC:** Oligodendrocyte-progenitor-like.

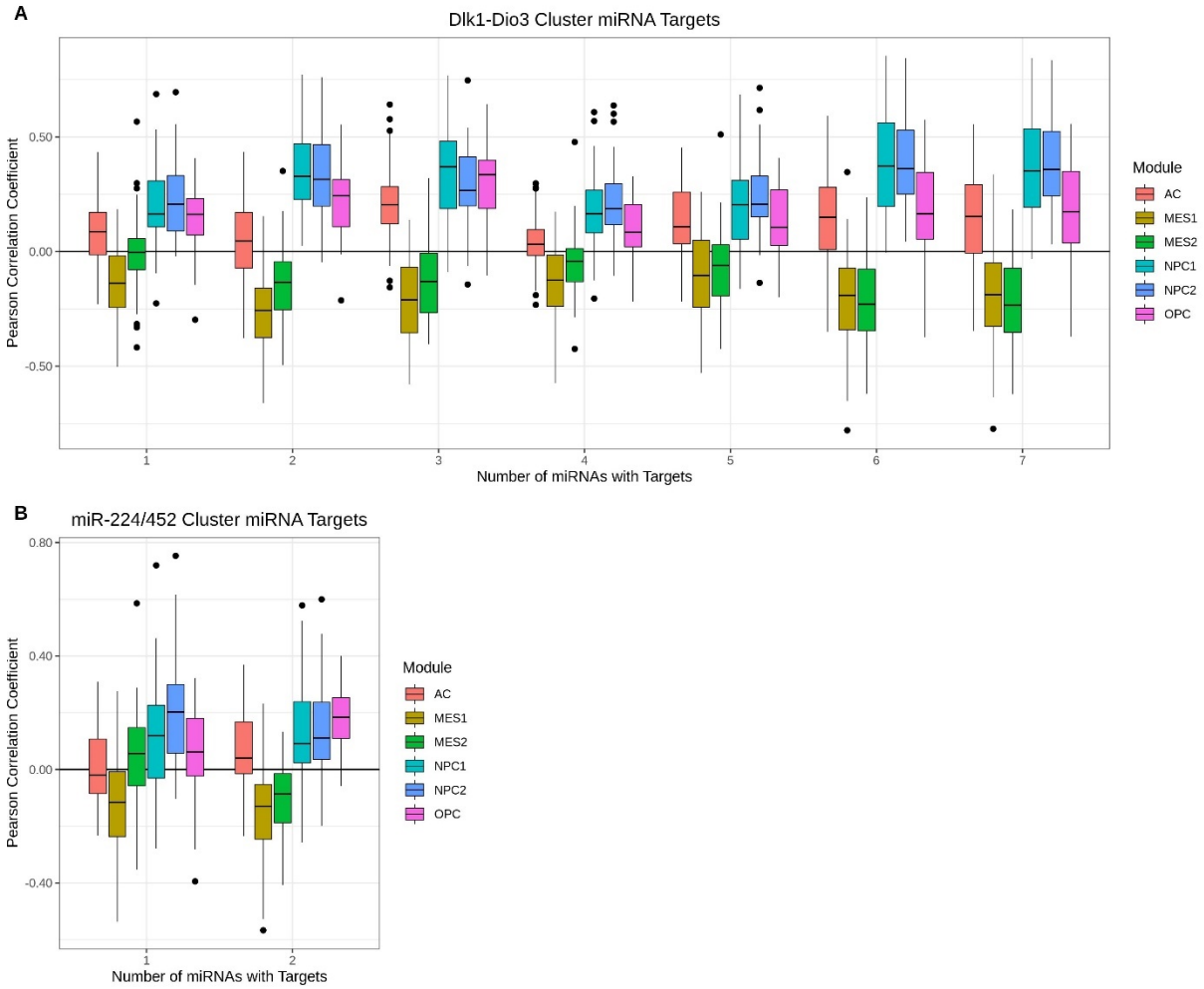

Figure 2.19. Correlation of scores for cluster targets and cell states in glioblastoma tumours. **A.** Multiple target scores were calculated for the Dlk1-Dio3 cluster by including genes which were common targets of up to 7 miRNAs. **B.** Two target scores were calculated for the miR-224/452 cluster, including all targets (1 minimum miRNA) and common targets of 2 miRNAs. Pearson correlations were calculated using cell scores from each tumour. **AC:** Astrocyte-like. **MES:** Mesenchymal-like. **NPC:** Neural-progenitor-like. **OPC:** Oligodendrocyte-progenitor-like.

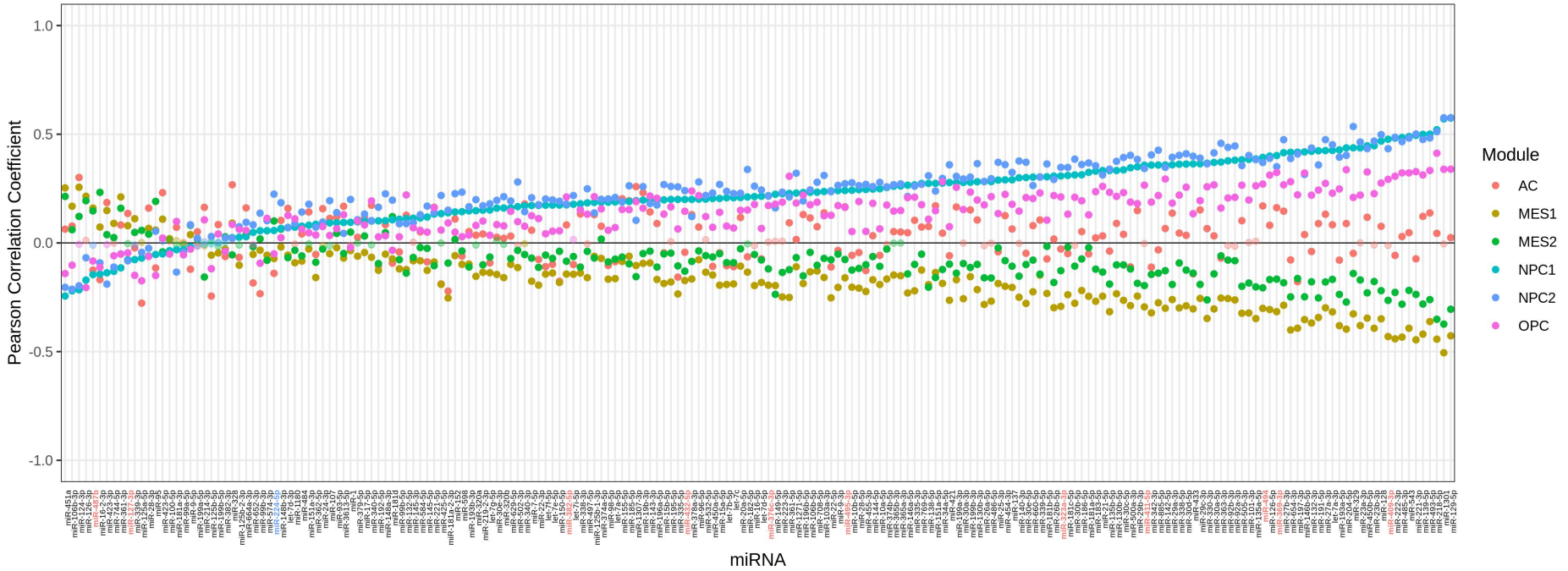

Figure 2.20. Correlation of miRNA target scores for 200 miRNAs highly expressed in glioblastoma tumours. miRNAs are ordered by ascending Pearson correlation coefficients for the NPC1 module. miRNAs are highlighted from the Dlk1-Dio3 locus (red) and miR-224/452 cluster (blue). Non-significant coefficients are shown as transparent dots. **AC:** Astrocyte-like. **MES:** Mesenchymal-like. **NPC:** Neural-progenitor-like. **OPC:** Oligodendrocyte-progenitor-like.

2.7. Discussion

In this chapter we investigated miRNA heterogeneity in glioblastoma to determine if miRNAs may play a role in the regulation of glioblastoma cell states. We identified two miRNA clusters in a primary glioblastoma culture which discriminated between two subpopulations of cells, suggesting they are expressed in a cell autonomous manner, and hypothesized that they may represent two different glioblastoma cell states. There were mixed results regarding both miRNA clusters and their association with the cell states defined in this study. The results suggest miRNA expression from the Dlk1-Dio3 locus, as well as the putatively co-expressed long non-coding RNA MEG3, is increased in NPC cells and decreased in MES cells. However, expression of their miRNA targets was also increased in NPC cells and decreased in MES cells, which suggests miRNA activity is higher in the MES cells and lower in the NPC cells. Several factors may have contributed to these observations, including an incomplete or oversimplistic characterization of glioblastoma cell states, poor performance of miRNA target prediction algorithms, or limitations with sequencing that challenge the inference of miRNA activity through expression of their targets. Furthermore, cause-and-effect is difficult to establish via association and it remains to be seen if any of these miRNAs are capable of driving or maintaining cancer cell states, or if their expression simply changes in response to other, more critical genes. Even in the absence of a functional role, miRNAs may still hold value as biomarkers for cell states once the field evolves towards clinical strategies which factor intra-tumoural heterogeneity.

2.7.1. Defining glioblastoma cell states

Although there is some evidence that discrete subpopulations of cells can be identified using RNA or miRNA expression, many large scale single cell studies have observed that cells are often continuously distributed between cell states with a high number of cells in 'intermediate' states[184,194,240,241]. Additionally, 'hybrid' states which have upregulated gene modules for multiple states have been detected in glioblastoma[184]. These 'hybrid' states do not appear completely random as some combinations were much more common than others. It is plausible that the models of glioblastoma cell states used in this study imperfectly capture the intra-tumoural relationships between glioblastoma cells and may have confounded the results.

A key question is whether cell states defined with RNA expression are also mirrored with changes in miRNA expression, or if examining each expression modality will lead to the detection of different subpopulations. For example, in a study with hepatocellular carcinoma cells, the authors used hierarchical clustering to identify 3 subpopulations of cells with RNA expression but found only 2 with miRNA expression[194]. There was significant overlap between 2 of the RNA expression-based subpopulations and the miRNA subpopulations, however the third was not identifiable in the miRNA

expression data. This suggests that it cannot always be assumed that there is concordance between RNA-based cell states and miRNA expression. Consequently, if the miRNAs examined in this study, including those from the Dlk1-Dio3 and miR-224/452 clusters, regulate cell subpopulations in a manner that is different to the cell states used in this study, then it is less likely that we would observe any strong association with them.

2.7.2. Challenges of determining miRNA activity through changes in target expression

In the absence of multi-omic sequencing data which captures miRNA and RNA expression in single cells, we opted to make use of more accessible scRNA-seq data for glioblastoma cells to infer miRNA activity through down-regulation of their targets. For target prediction we used miRNAtap, which aggregates target predictions from popular prediction tools (DIANA, Targetscan, PicTar, Miranda, and miRDB), and only used targets predicted by at least two tools to improve the detection of true positive targets[233,242–246]. However, target prediction still remains unreliable, and algorithms still fail to capture the true range of regulatory targets[247]. Early target prediction algorithms relied on the detection of canonical interactions where miRNA-target pairs were matched between the seed sequence and 3´UTR respectively[248]. More recent iterations of target prediction algorithms employ machine learning to improve performance and can also identify targets with non-canonical interactions[248,249]. Despite some improvements in performance over the years, experimental validation of targets are still needed to confidently identify regulatory interactions with miRNAs[247]. Unfortunately, the number of predicted miRNA-target interactions make it impractical to validate all miRNA-target pairs, though high throughput validation methods using CLIP or bind-n-seq approaches can help to alleviate this issue[250,251].

Measuring the biological impact of individual miRNAs is extremely challenging, and many studies fail to identify quantifiable relationships between expression of miRNAs and their expected targets. Modulating RNA levels is just one of several mechanisms that miRNAs use to regulate their targets, and other mechanisms such as translational inhibition are not detectable with this approach. Additionally, miRNAs are each predicted to regulate hundreds to thousands of genes and their influence on each gene may be minor, acting in concert with other miRNAs to be effective. Previous studies which examine mRNA expression profiles from microarrays or sequencing have shown that miRNA activity can be extrapolated from this data, indeed many target prediction algorithms such as miRDB and MirTarget are trained with mRNA expression data[246,252,253]. However, the suitability of single cell sequencing data for predicting miRNA activity has not been validated. Individual cells are inherently stochastic, and current methods for both single cell RNA and miRNA sequencing are plagued by technical issues which lead to noisy data and feature dropouts[254].

Finally, our work assumes that miRNA activity can be captured through negative correlations between miRNAs and the genes they regulate, however this may not always be true. For example, gene networks which involve co-expression of miRNAs and their targets, could lead to a positive correlation between the two despite the miRNAs role in suppressing these targets. This may also explain why NFIB, a marker gene of the NPC1 module, was targeted by 7 of the Dlk1-Dio3 miRNAs despite evidence that these miRNAs were upregulated in this cell state. When we scaled our analysis to include the targets of 200 miRNAs, this highlighted a broader trend of positive correlations with NPC and OPC scores and negative correlations with MES scores, indicating the correlations observed for the Dlk1-Dio3 and miR-224/452 miRNA clusters were not isolated to them. While it is possible that the miRNA pathway has a more generalized role in promoting or suppressing specific cell states, an alternative explanation is that the results are capturing correlations of cell state scores and gene expression in a manner unrelated to the miRNAs. For example, as sequencing is inherently relative, an increase in expression of RNAs which dominate one cell state may decrease measurements for the remaining RNAs, translating into a negative correlation with most mRNAs as that state is more highly expressed. Furthermore, as miRNA targets would be biased towards protein coding RNAs, a general change in their levels relative to other RNA classes within a cell state may also produce a false correlation. Thus, we were not able to determine if the miRNAs were truly having an impact on the expression of their targets.

2.7.3. miRNAs are potential biomarkers of glioblastoma heterogeneity

miRNAs such as miR-9-3p, miR-27a, and miR-23a, have previously been identified as markers for discriminating TCGA-PN and TCGA-MS subtypes. Interestingly, results from our study were consistent with this work, as these 3 miRNAs had some of the strongest correlation coefficients with the corresponding subtypes compared to other miRNAs, suggesting at least some the associations here are genuine. How well these observations translate to single cell level miRNA expression and their association with cell states remains to be confirmed.

Our study is the first to implicate the Dlk1-Dio3 and miR-224/452 miRNA clusters as potential regulators of glioblastoma heterogeneity. We identified an inverse relationship between RNA expression in cells with high expression of NPC/OPC genes compared to cells with high expression of MES genes. This observation was also evident in bulk miRNA expression data when comparing NPC/OPC and MES scores which suggests miRNA expression is highly responsive to cell states and likely incorporated into the gene networks regulating them. Although we have limited evidence that the two subpopulations of KS4 cells are most strongly associated with the NPC and MES cell states, further research which pairs miRNA sequencing with a more direct way of identifying biologically relevant cell states is critical. Furthermore, studies will need to be scaled up to included multiple

tumours to determine if heterogenous expression of miRNAs from the Dlk1-Dio3 and miR-224/452 clusters are a common feature in glioblastoma or an isolated case.

3. A comparative analysis of single cell small RNA sequencing data reveals heterogenous isomiR expression and regulation

3.1. Chapter introduction

miRNAs play a critical role in the regulation of gene expression in normal and cancerous cells. IsomiRs are variants of miRNAs with alterations in their nucleotide sequence, which can occur at the 5´ or 3´ ends of the miRNA where bases are added or subtracted, or anywhere within the sequence, including the seed region[255]. Variations at the 3´ end can also be distinguished in terms of whether they match their precursor RNA/originating gene sequence or not, as different sequences can be generated during miRNA biogenesis from proteins such as Drosha and Dicer, or by additional factors which can introduce new bases to the 3´ end. IsomiRs may possess alternative regulatory functions as evidence suggests that changes to bases at the 3´ end can alter a miRNAs regulatory activity or stability, whereas shortening or extending the 5´ end of a miRNA can affect target recognition due to a shift in the seed sequence[256]. Additionally, isomiR expression is closely associated with disease states, including cancer, however their role in many diseases is poorly understood.

Although most miRNA sequencing studies do not design experiments with a focus on studying isomiRs, isomiR research benefits from the fact that sequencing data captures complete nucleotide sequence information for small RNAs, and therefore isomiR abundance can also be estimated. To this date, miRNA and isomiR research has largely been based on population-level sequencing studies which are poorly equipped to investigate their role in intra-tumour heterogeneity[257]. Single cell small RNA sequencing methods which can capture and quantify miRNAs, and in principle isomiRs, have been recently developed but so far have not been utilized to their full potential[191,194,258]. Single cell data is challenging to work with and analysing single cell data for isomiR studies introduces additional problems which have not been addressed to this date[254]. A fundamental question that remains unanswered is whether isomiR expression is the result of stochastic processes or if they are under cell autonomous regulation.

Accurate and reliable isomiR quantification remains challenging as many of the issues with sequencing data analysis are more pronounced when studying isomiRs. As many studies have shown, isomiRs with as little as one nucleotide difference have potentially distinct regulatory functions[55,259]. Consequently, a common approach to isomiR quantification involves annotation and quantification of each unique sequence as distinct isomiRs[260–262]. However, it is well recognized that errors in read sequence are introduced in both the library preparation and sequencing steps, and this can lead to an overestimation of the number of unique isomiRs in a sample as well as an underestimation of the

quantity of certain isomiRs due to the mis-annotation of some reads[263,264]. This issue is expected to be more pronounced when working with single cell data, which is generally sequenced at a much lower read depth than bulk sequencing data, as sequencing costs are usually distributed across a larger number of samples. Additionally, the amplification step during library preparation introduces bias towards certain cDNAs and can lead to less accuracy in RNA quantification[265,266]. The low starting material of individual cells necessitates a higher number of PCR cycles, therefore greater bias, before reaching the required library size for sequencers and leads to a higher proportion of undesirable reads from RNA contamination. Several biochemical and bioinformatics approaches have been developed to address some of these issues. For example, error correction methods have been proposed to improve the accuracy of read sequences, although most have been designed for RNA-seq data and quantification. One error correction method was recently developed for miRNA analysis, using a k-mer lattice approach, however its efficacy on sequencing datasets with low read depth such as single cell data has not been demonstrated[267]. Unique molecular identifiers (UMIs) are short, randomized, artificial sequences that are introduced to cDNA prior to PCR amplification, which allows compensation for the bias introduced during this stage as reads derived from the same cDNA are expected to have identical sequences and UMI sequence[268]. UMIs have been utilized in a number of bulk and single cell sequencing methods including RNA-seq, small RNA-seq and ATAC-seq to improve quantification accuracy as well as DNA-seq to distinguish low frequency mutations and sequencing errors[261,269–272]. However, UMIs require a high sequencing depth to reliably distinguish real nucleotide changes from errors and concerns have been raised regarding the utility of UMIs in less complex libraries, such as small RNA sequencing libraries, where they may result in an underestimation of miRNA abundance due to excessive de-duplication[273]. Despite these concerns, single cell small RNA sequencing is currently our only tool to study cell-wide isomiR profiles and so more reliable methods of isomiR quantification are needed.

To our knowledge no study has used single cell small RNA sequencing to investigate isomiR expression or regulation on a single cell level[274]. In the following chapter we apply an alternative approach to isomiR quantification using biologically relevant isomiR types to overcome some of the issues from working with single cell data. We compare data from three single cell small RNA sequencing protocols and assess their strengths, pitfalls, or biases with respect to their isomiR quantification to inform future experimental designs. We also find evidence that some isomiR types exert different regulatory effects on their canonical targets, suggesting they are functional, and that isomiR regulation is cell autonomous. Together this work implicates isomiR expression and processing as an unexplored gene regulatory layer that potentially contributes to intra-tumour heterogeneity.

3.2. Specific methodology

3.2.1. Data Collection

Raw single cell small RNA sequencing data was obtained from the Gene Expression Omnibus (GEO) database under accession ids GSE81287 (glioblastoma and embryonic cells), GSE114071 (leukemia cells) and the Genome Sequence Archive (Beijing Institute of Genomics) under accession ID CRA001133 (hepatocellular carcinoma cells). Bulk small RNA sequencing data was obtained from accession id GSE141687 (leukemia), GSE166349 (hepatocellular carcinoma), and GSE76903 (also hepatocellular carcinoma).

3.2.2. Sequencing Data Processing

For samples derived from the Smallseq protocol, UMI sequences were removed prior to any adapter removal and appended to the read headers[188]. Adapters were then removed using cutadapt (v2.7) with a minimum overlap of 1 nt and maximum error rate of 0.1 between reads and adapter sequences[228]. After UMI and adapter removal, reads shorter than 15 nucleotides were excluded.

To identify duplicated reads, reads were aligned to the human genome (hg38) using bowtie (v1.2.3) with the following parameters: -n 2 -e 120 -l 20 --best[229]. Human-aligned reads were subsequently deduplicated with umitools (v1.0.0) with default settings[230].

For the other protocols, as well as the non-deduplicated Smallseq analysis, separate pipelines which excluded the alignment to the human genome and subsequent deduplication with umitools were used.

3.2.3. miRNA Mapping and Annotation

Processed reads were aligned to miRbase (v22.1) annotated precursor miRNAs using miraligner (v3.4), with the following parameters: -sub 1 -trim 3 -add 3[36,231]. Reads which successfully aligned to a miRNA were also annotated with any variants to the miRbase defined mature sequence.

The following isomiR categories were defined: Canonical – miRNAs with a perfect match to the miRbase mature sequence, 5p Variant – miRNAs differing at the 5` end with respect to the miRbase mature sequence, 3p Template – miRNAs deviating in length to the miRbase mature sequence but still matching the precursor miRNA sequence, 3p Non-template – miRNAs which did not match the precursor miRNA sequence at the 3` end, and Substitution – miRNAs containing a maximum of 1 mismatch to the miRbase mature sequence, excluding variations at the 5` or 3` ends. Categories were then assigned any alignments which contained their respective isomiR modification and were used for measuring their proportions relative to the total number of annotated reads.

3.2.4. Correlation of miRNA and isomiR Expression with Predicted Targets

For building miRNA predicted target lists, the miRNAtap (v1.20) package was used to aggregate targets across the five supported algorithms (DIANA, Miranda, PicTar, TargetScan and miRDB) using the 'minimum' method and only considering targets predicted by 2 or more algorithms[233]. The Pearson correlation between miRNA and gene expression across all cells was calculated and used for plotting each miRNA with its targets.

3.2.5. Code for Data Analysis and Figures

Documents containing code used to generate the results in this section can be found in the following link: https://cloudstor.aarnet.edu.au/plus/s/GaNTw7rMWztjliu

3.3. miRNA and IsomiR abundance are highly variable across cell types in the three single cell small RNA-seq protocols

IsomiRs have been successfully used to classify cancer types and may have a role to play in oncogenesis[142]. However, so far isomiRs have not been studied on a single cell level and the mechanisms behind isomiR expression and processing across and within cell types are poorly understood.

To assess miRNA and isomiR expression in different cells and single cell sequencing protocols we analysed 9 cell types from 3 different studies (Figure 3.1 and Table 3.1). This included the seven cell types sequenced in the Small-seq study[188], which after quality control and filtering included 139 cells from three glioblastoma primary cell cultures, 35 cells from the U87 glioblastoma cell line, 48 cells from the human embryonic kidney cell line HEK293FT, and 107 naïve and 95 primed human embryonic stem cells (Table 3.1). Additionally, we included the 19 K562 leukemia cells from Wang et al's miRNA/mRNA Co-sequencing study (Co-seq)[191] and 32 hepatocellular carcinoma (HCC) cells isolated from a resected tumour sample and sequenced in the Holo-seq study[194]. Using miRNA expression, we calculated the Spearman's correlation between all pairs of cells and clustered them using hierarchical clustering (Figure 3.1A). We found that individual cells from the same cell type typically clustered together, except for several of the glioblastoma cell types which formed mixed clusters likely owing to their similar miRNA profiles. Using the same methodology for read mapping and miRNA annotation for all samples, we found that cells from the Small-seq protocol had a higher number of unique miRNAs (Figure 3.1B). This was despite a larger number of total reads being sequenced from the Co-seq and Holo-seq protocols (Table 3.1). However, the cells from the Co-seq and Holo-seq protocols had a higher number of unique isomiRs than most of the other cell types in the Small-seq protocol (Figure 3.1C).

Cell types from each protocol had characteristic miRNA lengths (Figure 3.1D) and expression levels (Figure 3.2). Most of the cell types from the Small-seq study had strong peaks at 22 nt, except for the primed embryonic stem cells which had a high number of 22nt and 23nt miRNAs. A significant proportion of miRNAs, averaged across cell types, in both the Small-seq glioblastoma (42.7-53.7%) and embryonic stem (37.9-52.6%) cell types were annotated as canonical. This was in stark contrast to the Holo-seq and Co-seq protocols where the majority of miRNAs in the HCC and K562 cells were isomiRs, with canonical miRNAs only representing 5.9% and 5.6% of all miRNAs respectively. It is worth noting that miRNA expression for some of the cell types, including the HCC and K562 cells, were dominated by a small number of miRNAs which would have skewed the overall lengths in favour of those miRNA profiles (Figure 3.2). As the HCC and K562 cell types were sequenced using different protocols from

independent labs, it is difficult to know how well the significant differences in isomiR abundance reflect cell specific differences in miRNA processing, maturation, or turnover, and how much is due to technical reasons such as experimental artifacts or bias. Notably, when we compared the single cell data to bulk RNA-seq data from other studies (Figures 3.3, 3.4, and 3.5), we found the bulk data had length distributions which resembled a more typical 22-23nt peak in both HCC tumour samples and K562 cells as opposed to the dual peaks observed in the HCC and K562 single cell data (Figure 3.3). Additionally, there was a higher proportion of canonical miRNAs in the bulk sequencing data (Figure 3.4 and 3.5).

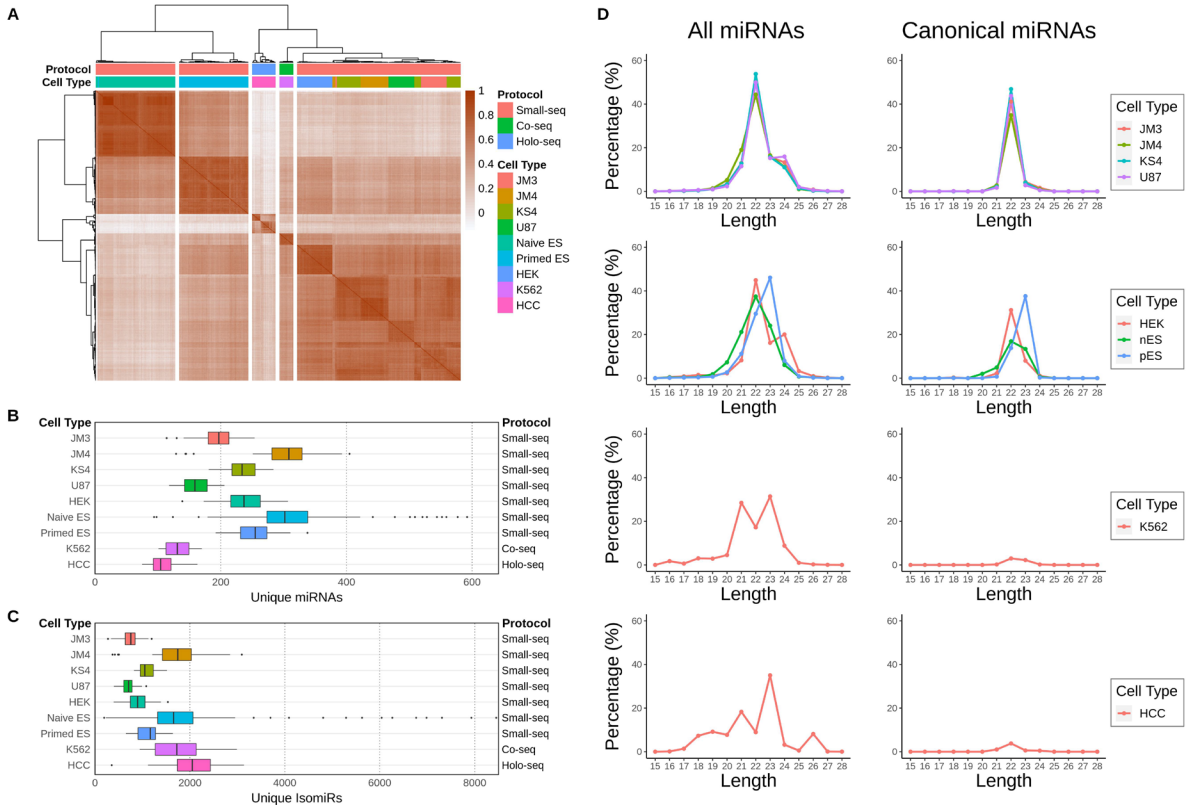

Figure 3.1. An overview of the small RNA sequencing samples that were analysed. **A.** Heatmap showing hierarchical clustering using Spearman correlation of miRNA expression between all samples. **B.** Number of unique miRNAs detected in each cell type. **C.** Number of unique isomiRs detected in each cell type. **D.** Analysis of total miRNA and isomiR length distribution profiles. Distribution of miRNA lengths considering all miRNAs (left) and canonical miRNAs only (right; according to miRbase). For canonical miRNAs, percentages are relative to total miRNA reads. Includes glioblastoma cell lines JM3, JM4, KS4, U87 (Small-seq protocol), embryonic kidney cells (HEK293FT), naïve embryonic (nES), primed embryonic stem cells (pES; Small-seq protocol), K562 cells (Co-seq protocol) and hepatocellular carcinoma cells (HCC; Holo-seq protocol).

Cell Type	Cell Origin	Culture Type/Source	Study/Protocol	No. of Cells Included (this study)	Average No. Reads Mapped to miRNAs (% of Total Reads)	Avg No. UMIs Mapped to miRNAs	Matched RNA-seq data
JM3	Glioblastoma	Primary Cell Culture	Faridani et al[188] (Small-seq)	37	6576.46 (0.21%)	2607.30	No
JM4	Glioblastoma	Primary Cell Culture	Faridani et al[188] (Small-seq)	42	27847.19 (0.67%)	7064.76	No
KS4	Glioblastoma	Primary Cell Culture	Faridani et al[188] (Small-seq)	60	14208.52 (0.41%)	4218.68	No
U87	Glioblastoma	Primary Cell Culture	Faridani et al[188] (Small-seq)	35	7819.03 (0.18%)	2321.09	No
Naïve Embryonic Stem Cells	Embryonic Stem Cells	Cell Line	Faridani et al[188] (Small-seq)	107	43382.87 (1.03%)	6223.42	No
Primed Embryonic Stem Cells	Embryonic Stem Cells	Cell Line	Faridani et al[188] (Small-seq)	95	20866.96 (0.56%)	3319.39	No
HEK293FT	Embryonic Kidney Cells	Cell Line	Faridani et al[188] (Small-seq)	48	8390.38 (0.24%)	2908.21	No
K562	Leukemia	Cell Line	Wang et al[191] (Co-seq)	19	25090.95 (0.38%)	N/A	Yes
Hepatocellular Carcinoma (HCC)	Hepatocellular Carcinoma	Tumor	Xiao et al[194] (Holo-seq)	32	31651.66 (0.38%)	N/A	Yes

Table 3.1. General features of the single cell small RNA datasets. Number of cells and average mapping values are according to the mapping and filtering methodology used in this study.

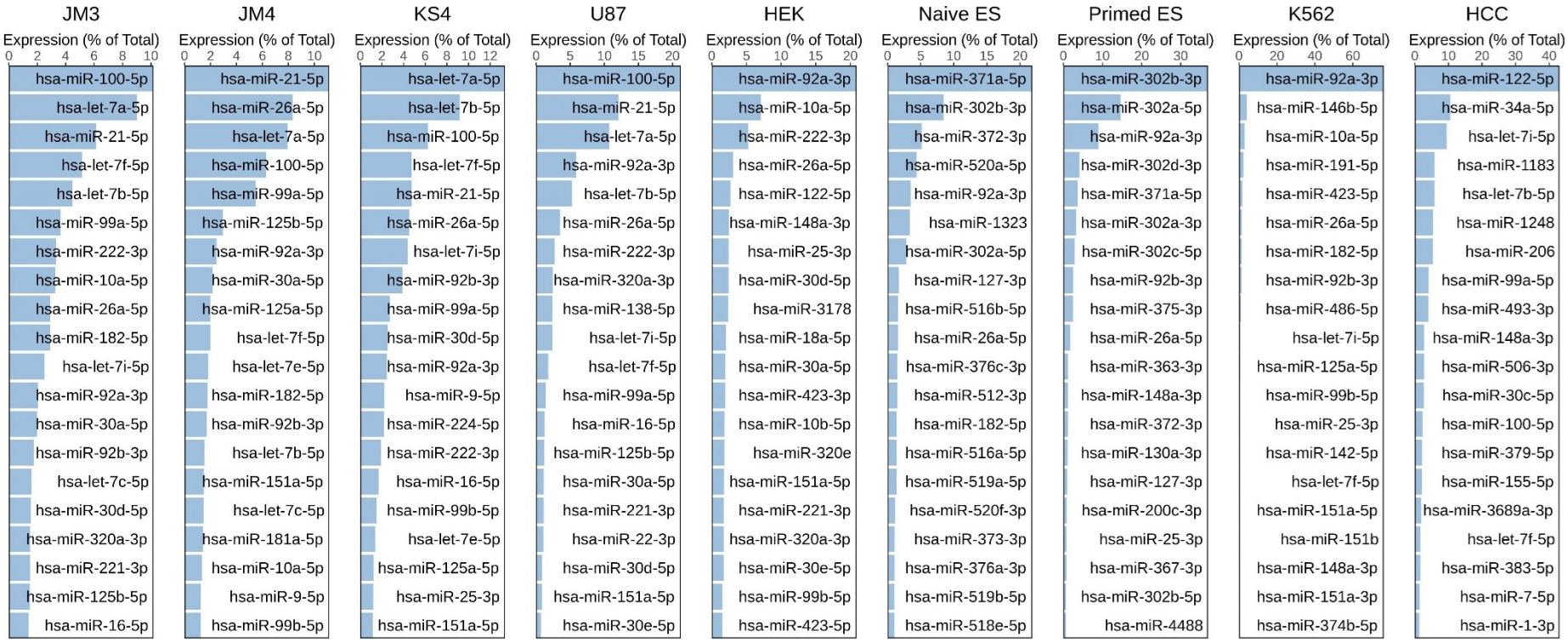

Figure 3.2. Top 20 miRNAs by expression for each cell type included in this study. Percentages shown are in proportion to total reads mapped to miRNAs. Includes the glioblastoma primary cultures/cell lines JM3, JM4, KS4 and U87, human embryonic kidney cell line HEK293 (HEK), naïve and primed embryonic stem (ES) cells, the K562 leukemia cell line and hepatocellular carcinoma cells (HCC).

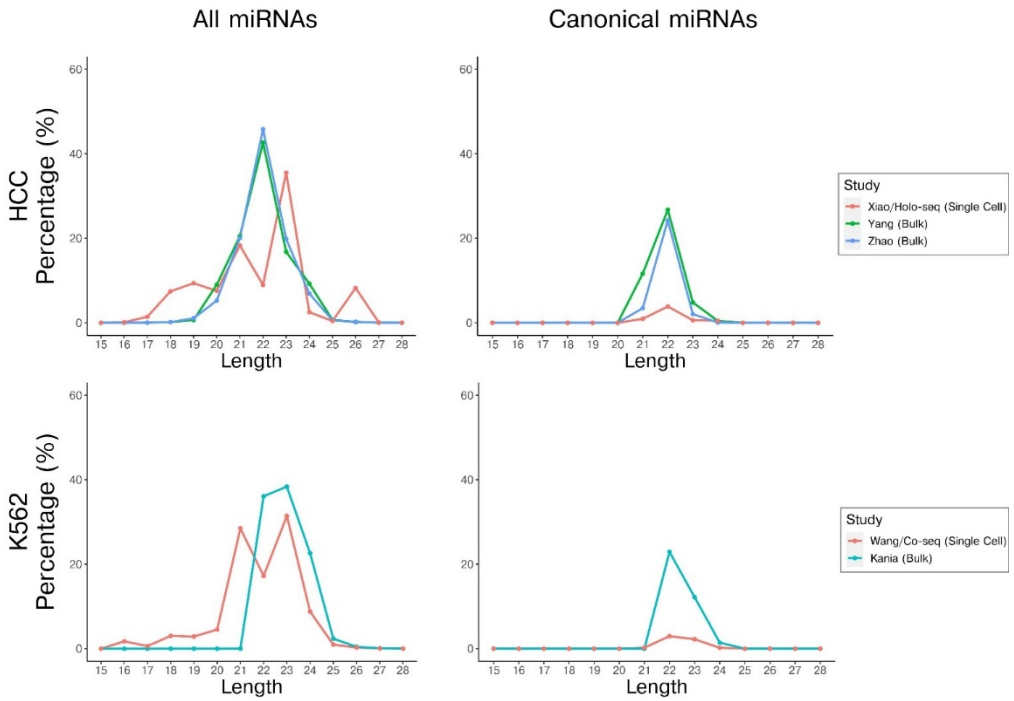

Figure 3.3. Comparison of total miRNA and isomiR length distribution profiles between single cell and bulk RNA-seq datasets from independent studies. Includes HCC tumour and K562 cells. Distribution of miRNA lengths considering all miRNAs (left) and canonical miRNAs only (right; according to miRbase). For canonical miRNAs, percentages are relative to total miRNA reads.

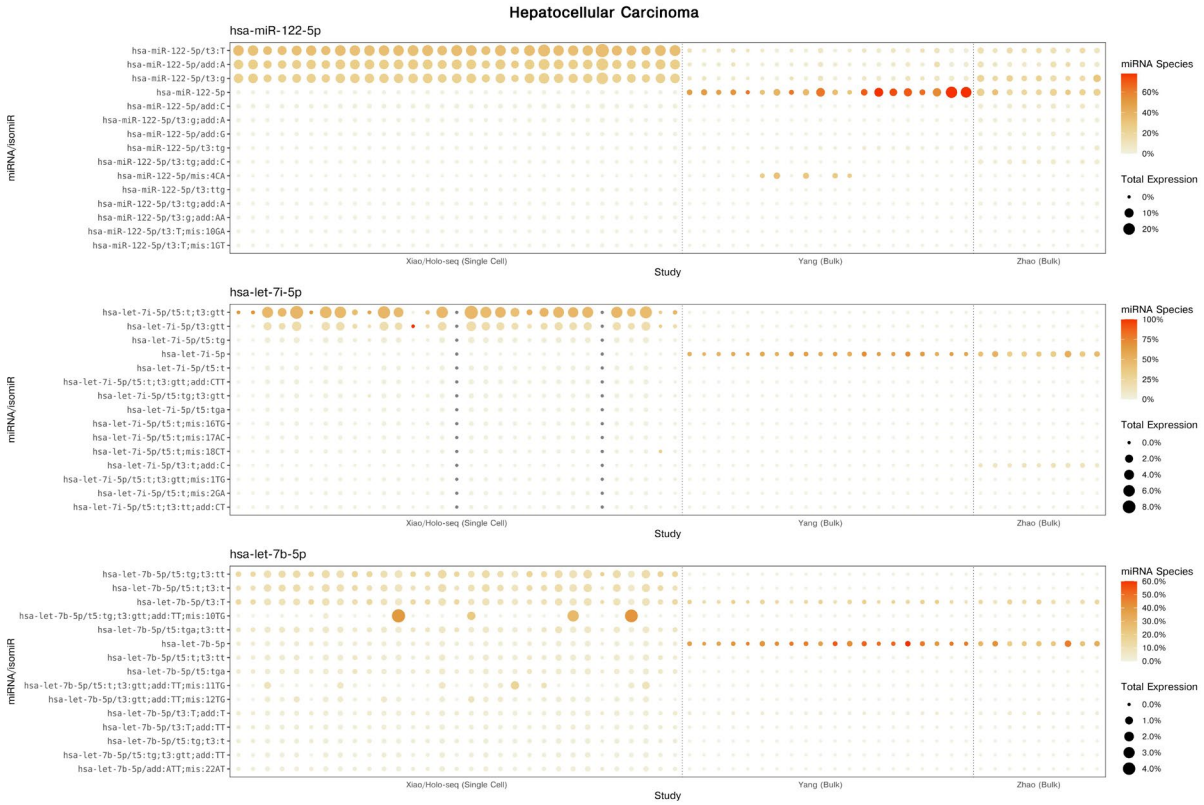

Figure 3.4. Expression and relative abundance of isomiRs in Hepatocellular Carcinoma single cell and bulk smRNA-seq data from independent studies. Shows each isomiRs expression (rows) across individual cells (columns), with respect to their relative abundance of their miRNA gene (dot colour) and the isomiRs expression normalized to total miRNAs for each cell (dot size). Top 15 isomiRs for the 3 highest expressed miRNAs in the single cell dataset are displayed. Cells which did not have any reads mapping to the miRNA are indicated by grey circles.

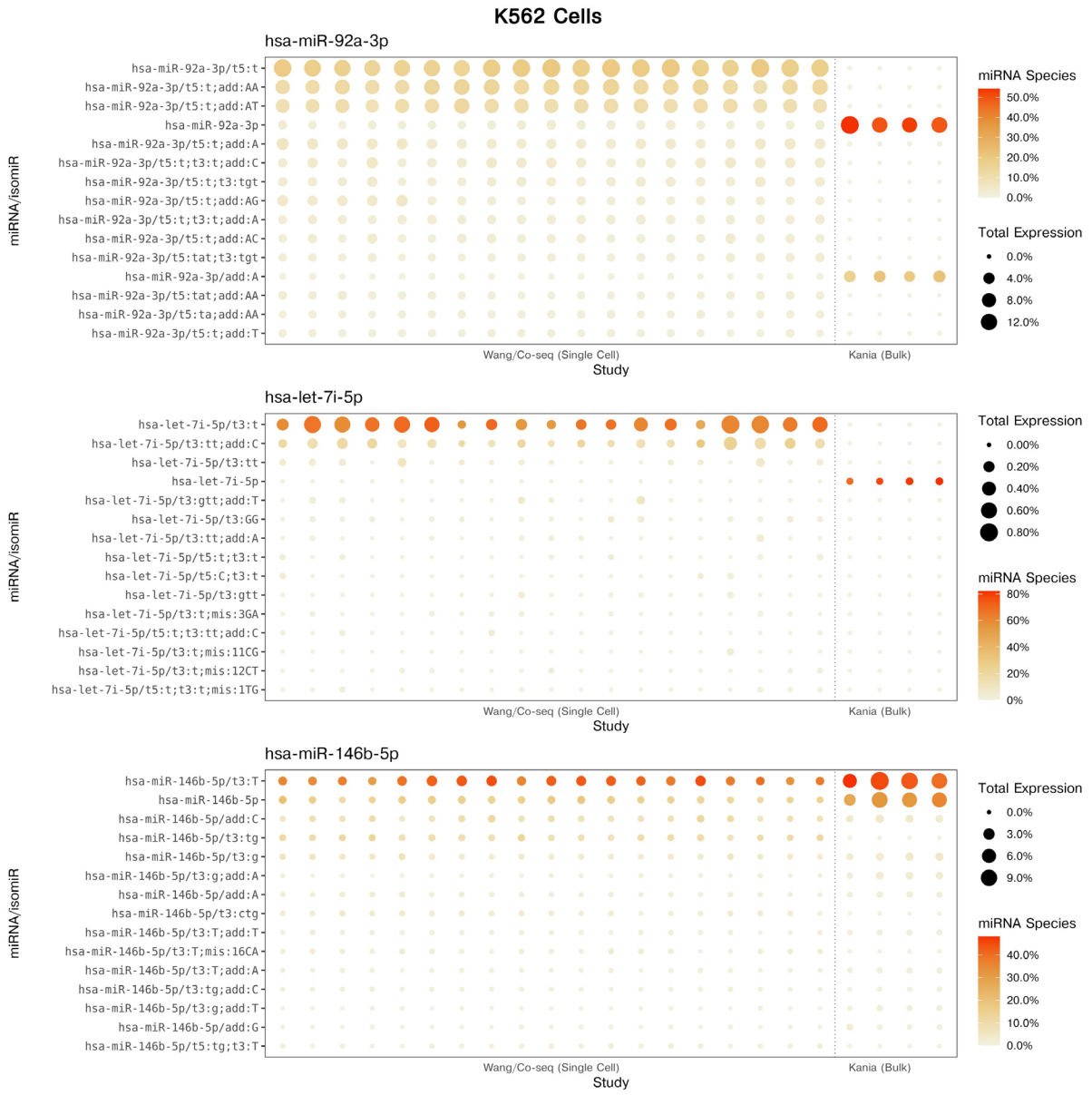

Figure 3.5. Expression and relative abundance of isomiRs in K562 Leukemia single cell and bulk smRNA-seq data from independent studies. Shows each isomiRs expression (rows) across individual cells (columns), with respect to their relative abundance of their miRNA gene (dot colour) and the isomiRs expression normalized to total miRNAs for each cell (dot size). Top 15 isomiRs for 3 highly expressed miRNAs in the single cell dataset are displayed. Cells which did not have any reads mapping to the miRNA are indicated by grey circles.

3.4. IsomiR processing is likely regulated by cell autonomous mechanisms

3.4.1. Global expression of isomiRs is specific to cell types

In the previous section we found substantial differences in miRNA and isomiR abundance between cell types. Following this we compared isomiR processing between individual cells by focusing on the 5′ and 3′ templated positions in each miRNA, calculating the proportion of isomiRs (averaged across single cells in each cell type) starting or ending at each position +/- 3 nucleotides (nts) around the canonical ends (Figure 3.6A). Additionally, we included the isomiRs with adenine or uridine bases added to the 3′ end (Figure 3.6A) as many studies have shown isomiRs with these additions are more common and can lead to changes in miRNA stability and target recognition[56,275,276].

5′ variants which extended the isomiR from the canonical site were almost completely absent in all cells (Figure 3.6A). 5′ variants shorter than the canonical site were also rare, particularly in the Small-seq derived cells, except for primed embryonic stem cells which had a notable amount of isomiRs shortened at the 5′ end by 3 nucleotides (13.0%). However, for the K562 cells, there was a high number of 5′ variants predominantly 1 nt shorter than the canonical site (61.7%), with smaller amounts of 2 and 3 nt shortened isomiRs (4.7% and 8.7% respectively).

With 3′ variants, all cells from the Small-seq and Co-seq protocols showed very similar 3′ templated isomiR profiles according to position (Figure 3.6A), with a strong peak at the canonical site and sharp decline in proportions for the surrounding 1-2 nts. However, HCC cells had nearly equivalent abundances of miRNAs with 3′ canonical sites as well as those shortened or extended by 1 nucleotide, indicating more extensive processing at the 3′ end. Non-templated 3′ variants containing adenine and/or uridine were present at low levels (12.8-24.2%) in all cell types except K562 cells, where more than half the isomiRs had an untemplated addition (58.2%), most with at least one non-templated adenine (48.8%).

To investigate potential cell-specific mechanisms involved in isomiR expression and processing we labelled isomiRs according to several categories that describe sequence alterations likely to be driven by distinct mechanisms[256]. We then calculated the proportions of every cell's miRNA possessing each isomiR alteration (Figure 3.6B). Here, we defined the following isomiR types: canonical (identical sequence to the miRbase annotated miRNA), 5′ variants (altered sequence on the 5′ end, still matching the precursor miRNA sequence), 3′ templated (altered 3′ sequence, but still matching the precursor sequence), 3′ non-templated (altered 3′ sequence with bases not matching its precursor), or substitutions (single nucleotide differences compared to the canonical sequence, excluding 5′ and 3′ ends). Note that each isomiR may contain several sequence alterations and therefore contribute to multiple isomiR types.

Cells from the Small-seq protocol generally showed similar proportions of isomiR types, with 3′ templated variants (34.8-50.9%) being the dominant isomiR type (Figure 3.6B). Minor differences were observed between cell types which were statistically significant for nearly all isomiR types in this protocol. The majority of miRNAs from the Co-seq K562 cells were 5′ variants (76.1%), with a high proportion of 3′ non-templated variants (58.2%) and similar levels of 3′ templated variants (34.2%) to the Small-seq cells. On the other hand, HCC cells from the Holo-seq protocol produced predominantly 3′ templated (67.3%) and 5′ variants (27.3%), and there was a large increase in variability among single cells across all categories which may reflect the heterogeneity of cells sourced from a tumour sample compared to cell lines or primary cell cultures.

3.4.2. Processing of isomiRs is unique to each miRNA's gene

We then re-analysed isomiRs by their 5′ and 3′ templated locations and adenine/uridine additions, but this time separating isomiRs by their miRNA gene (Figure 3.7) and found that different miRNAs exhibited unique patterns of isomiR expression both at the 5′ and 3′ ends. Expression levels of 3′ templated variants were clearly different between miRNAs from the same cell type (Figure 3.7). We did not identify high levels of 5′ variants in any of the Small-seq cells (glioblastoma or embryonic stem cells) when looking at highly expressed miRNAs (Figure 3.7). While non-canonical 5′ variants were pervasive in nearly all miRNAs in HCC cells, K562 cells showed a mixture of some miRNAs being highly variable at the 5′ end and others not at all, suggesting miRNA-specific processing at this end.

3′ non-templated additions were generally rare in the most highly expressed miRNAs (Figure 3.7). One exception was miR-92a-3p, where in K562 cells, 60.8% of miR-92-3p had at least one non-templated adenine and 20.9% with at least one uridine. This coincided with a high overall expression of the miRNA (76.3%) and suggests adenine additions may contribute to the stabilization of miR-92a-3p in K562 cells which would be consistent with previous studies[56]. Adenine and relatively lower levels of uridine additions were observed in all cell types for miR-92a-3p, albeit to a lesser extent than K562 cells.

Overall, we observed that the profiles of the relative abundance of different 5′ and 3′ variants were unique to each miRNA, and it appears that both 5′ and 3′ processing may be driven by mechanisms that distinguish between miRNA species.

3.4.3. Individual cells exhibit cell type specific isomiR processing

As several of the miRNA genes were expressed across multiple cell types, we compared how miRNAs were processed into isomiRs in different biological contexts (Figure 3.7 and 3.8). In general, single cells from the same cell type were highly similar with respect to their isomiR processing at both the 5′ and 3′ ends (Figure 3.7). However, there were substantial differences observed between cells from some

of the cell types. This was most clear when comparing cell types from the 3 different protocols but was also evident to a lesser extent when considering cell types from the Small-seq protocol. For example, when comparing isomiR processing of let-7a-5p (Figure 3.7A), cell types from the Small-seq protocol expressed isomiRs which predominantly matched the canonical sequence (position 0) at both the 5′ and 3′ ends when considering their templated sequence. However, there were exceptions with the K562 cells which expressed a high proportion of shortened 3′ templated isomiRs, as well as the HCC cells which expressed a high proportion of shortened 5′ templated isomiRs. With let-7b-5p, the HCC cells had high expression of non-canonical processed isomiRs at both the 5′ and 3′ ends (Figure 3.7B). This contrasted with the cell types from the Small-seq protocol which were predominantly canonical at both their 5′ and 3′ templated ends (Figure 3.7B). There was also an increase in variation between some of the cell types from the Small-seq protocol, including the HEK and Naïve Embryonic Stem cells, that was not evident in the glioblastoma cell types (JM3, JM4, KS4, and U87). For miR-125a-5p (Figure 3.7C), 3′ templated isomiR expression was skewed to favour the canonical location for some cells such as JM3 and HEK but shorter variants were dominant in JM4 and K562 cells (Figure 3.7C).

Overall, the data supports a cell autonomous isomiR regulatory system as each cell type has an intrinsically determined isomiR expression level and individual cells can regulate their processing of isomiRs with a high degree of precision. The major differences between cell types in relative isomiR abundance of certain miRNA species suggests cell autonomous mechanisms may be involved in miRNA processing and isomiR expression.

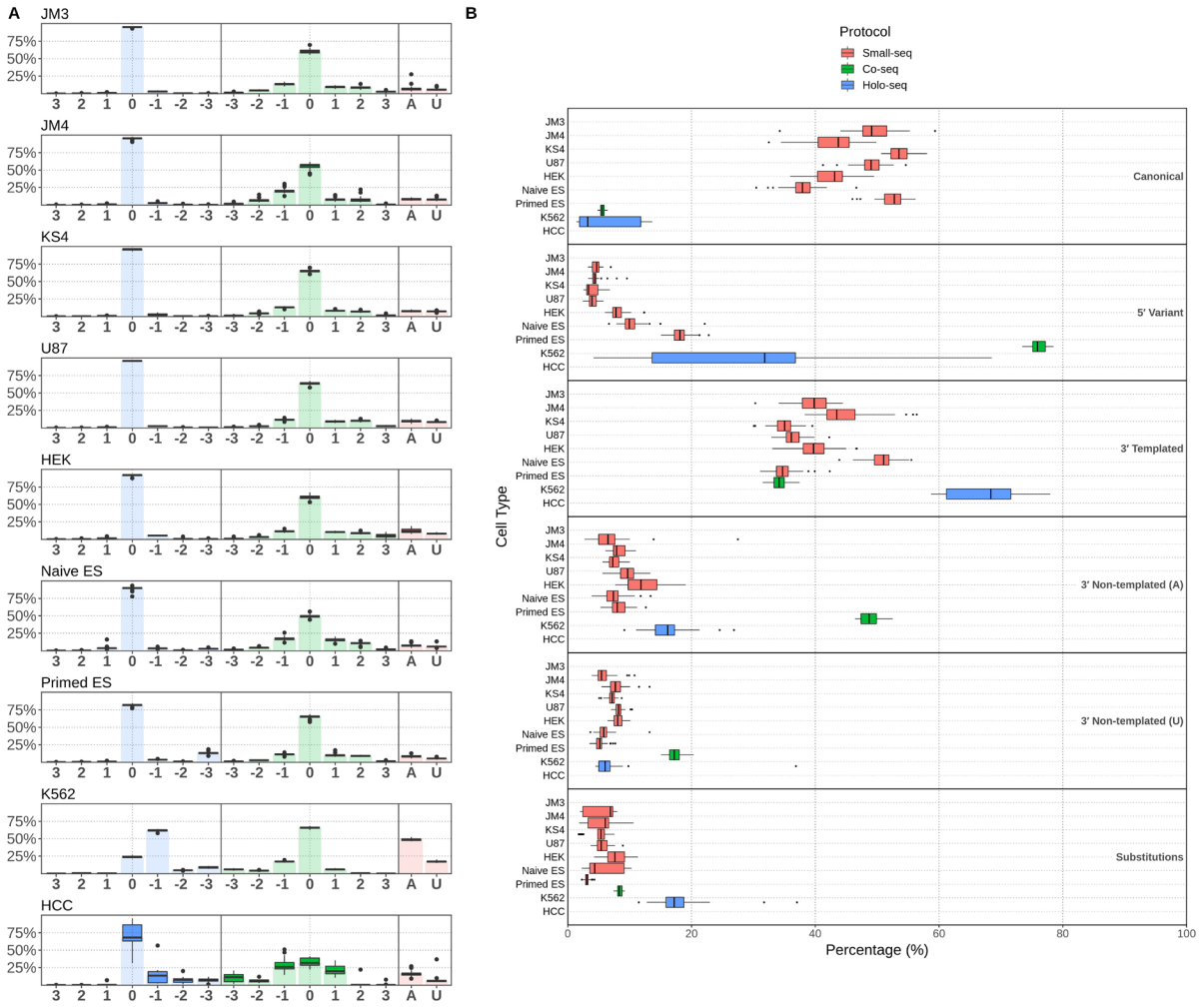

Figure 3.6. Comparison of relative isomiR expression in each cell type. **A.** Shows the relative templated proportion of miRNAs according to their 5' end (blue) or 3' end (green) templated (T) locations, as well as proportions of miRNAs with non-templated (UT) Adenine (red; A) or Uridine (red; U) additions. **B.** Box plots constructed from single cell data, showing percentage of total miRNAs belonging to each isomiR category. **ES:** Embryonic Stem Cell.

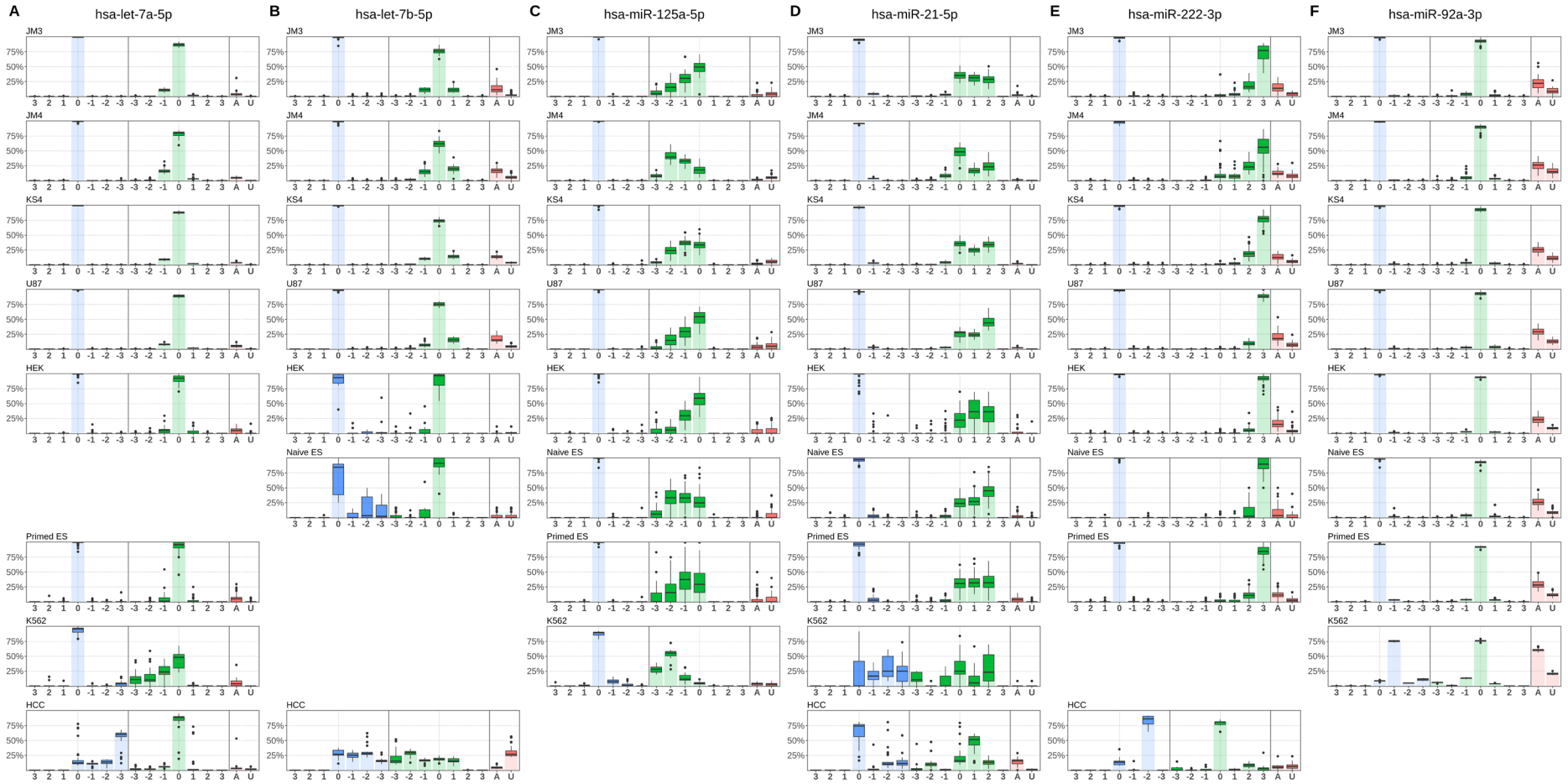

Figure 3.7. Comparison of 5′ and 3′ isomiRs in select miRNAs that are highly expressed across multiple cell types. Shows the relative templated nucleotide position of miRNAs (averaged) at the 5′end (blue) or 3′ end (green), with respect to the canonical miRNA (according to miRbase). Includes **A.** let-7a-5p, **B.** let-7b-5p, **C.** miR-125a-5p, **D.** miR-21-5p, **E.** miR-222-3p, and **F.** miR-92a-3p. Non-templated additions for Adenine and Uridine are also shown (red). miRNAs with absent plots were either not expressed or did not pass filter criteria. **3′ UT:** 3′ Non-templated addition. **ES:** Embryonic Stem Cell.

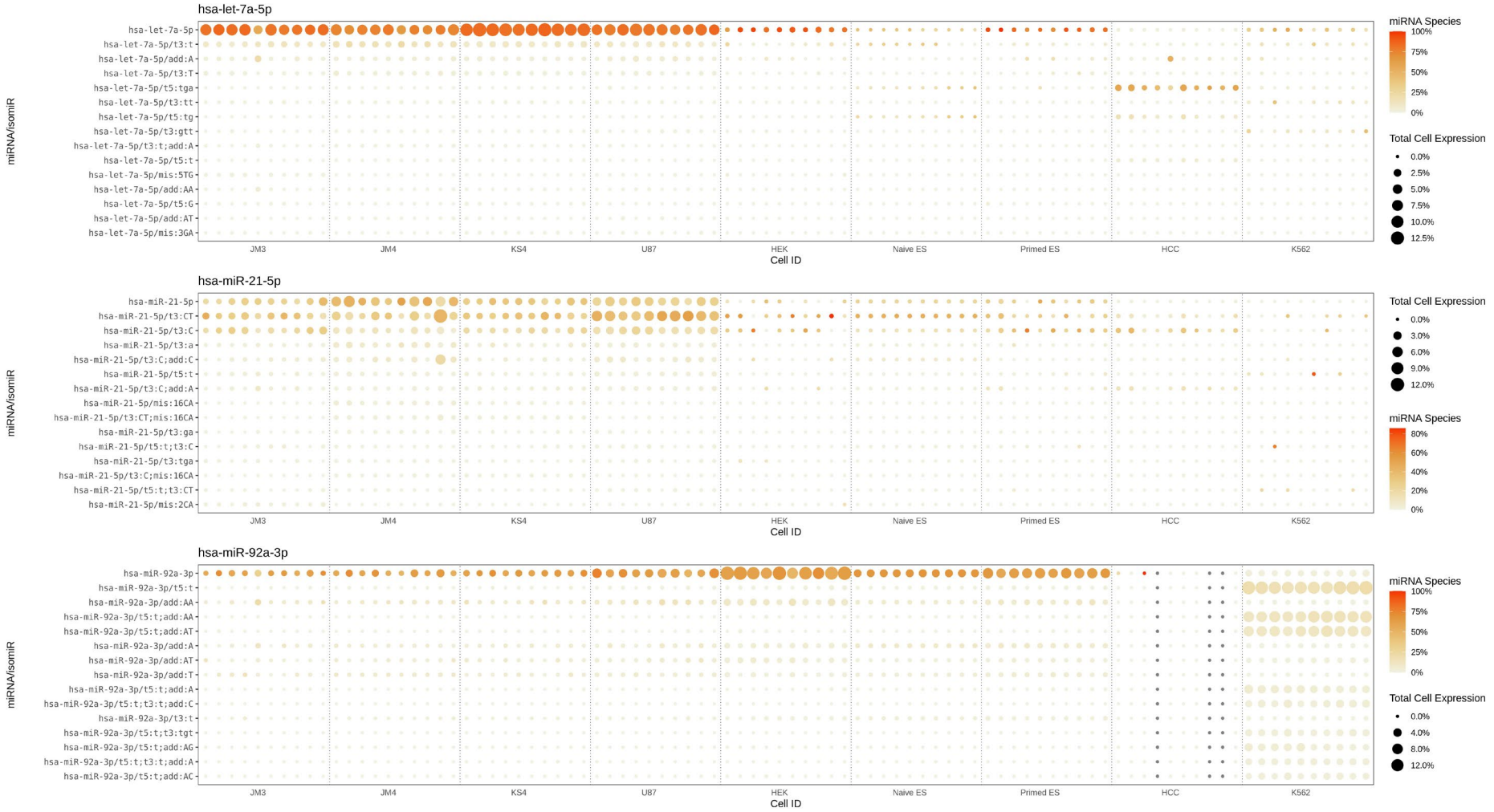

Figure 3.8. Expression and relative abundance of isomiRs in single cells. Shows each isomiRs expression (rows) across individual cells (columns), with respect to their relative abundance of their miRNA gene (dot colour) and the isomiRs expression normalized to total miRNAs for each cell (dot size). Top 15 expressed isomiRs for let-7a-5p, miR-21-5p and miR-92a-3p are displayed. 10 cells from each cell type are included. Cells which did not have any reads mapping to the miRNA are indicated by grey circles.

3.5. IsomiR processing alters regulation of their canonical targets

The combination of miRNA and mRNA expression data available for the K562 and HCC cells provided us with an opportunity to estimate how different isomiR types could affect regulation of their predicted canonical targets. For this, we generated predicted target lists for the most highly expressed miRNAs using miRNAtap, a tool which aggregates target predictions from multiple sources[233]. Using expression data from each cell, miRNA expression was correlated against expression of each of the miRNAs predicted targets. As a negative control, correlation between each miRNA's expression and expression of each transcript not predicted to be targeted by the respective miRNA was determined. For each miRNA we also calculated the correlation between expression of isomiRs in each isomiR category and the miRNAs predicted canonical targets, calculated by taking the normalized count of reads mapping to a given miRNA that contained that particular isomiR type. IsomiR categories included – Canonical, 5′ Variant, 3′ Templated, 3′ Non-templated adenine (A), 3′ Non-templated uridine (U) and Substitutions. Results were displayed as a cumulative distribution of all miRNA-gene correlations (Figure 3.9 and 3.10). A higher proportion of negatively correlated genes in the target gene set compared to the negative control (non-targets) would cause the line to shift to the left and would suggest that the miRNA was effectively downregulating its targets.

We observed that the relationship between miRNA expression and expression of its predicted canonical targets varied between miRNAs in both cell types (Figure 3.9 and 3.10). As was shown in the original Co-seq study, in K562 cells (Figure 3.9) there was a stronger negative correlation of the dominantly expressed miR-92a-3p with predicted canonical targets (total) when compared to its correlation with all other transcripts (non-targets)[191]. In contrast, miR-182-5p was more positively correlated with predicted canonical targets (total) compared to non-targets, and other highly expressed miRNAs such as miR-146b-5p, miR-10a-5p, miR-191-5p, and miR-423-5p were not found to be significantly different. In the HCC cells (Figure 3.10), several highly expressed miRNAs were more negatively correlated with their predicted targets compared to non-targets, including miR-34a-5p and miR-148a-3p. Conversely, miR-122-5p, let-7i-5p, and let-7b-5p were more positively correlated, and no significant difference was found with miR-206. The results suggest that despite high expression in these cells, many miRNAs may exert a weak or absent regulatory effect on the expression levels of their predicted canonical targets. Results from the two-sided Kolmogorov–Smirnov test and read counts for isomiR-target pairs can be found in Table Appendix B1 and Appendix B2.

When analysing correlations between expression of canonical miRNAs or each isomiR category with the expression level of the miRNA's predicted canonical targets it was clear that there were differences in certain isomiR categories for several miRNAs. This was most evident in the K562 cell line where

canonical miR-92a-3p and its isomiRs with 5′ Variant or 3′ Templated changes had a negative correlation with its predicted targets. Contrasting this, miR-92a-3p with Non-templated (Adenine) or Non-templated (Uridine) additions were positively correlated (Figure 3.9). Interestingly, with miR-146b-5p, 5′ variants had the strongest negative correlation with predicted canonical targets compared to the other isomiR categories, including canonical miRNAs which was not significantly different to non-targets. We did not observe any consistency in the shift in correlations of isomiR categories when comparing different miRNAs, indicating that the effect of isomiR types on gene regulation is miRNA specific.

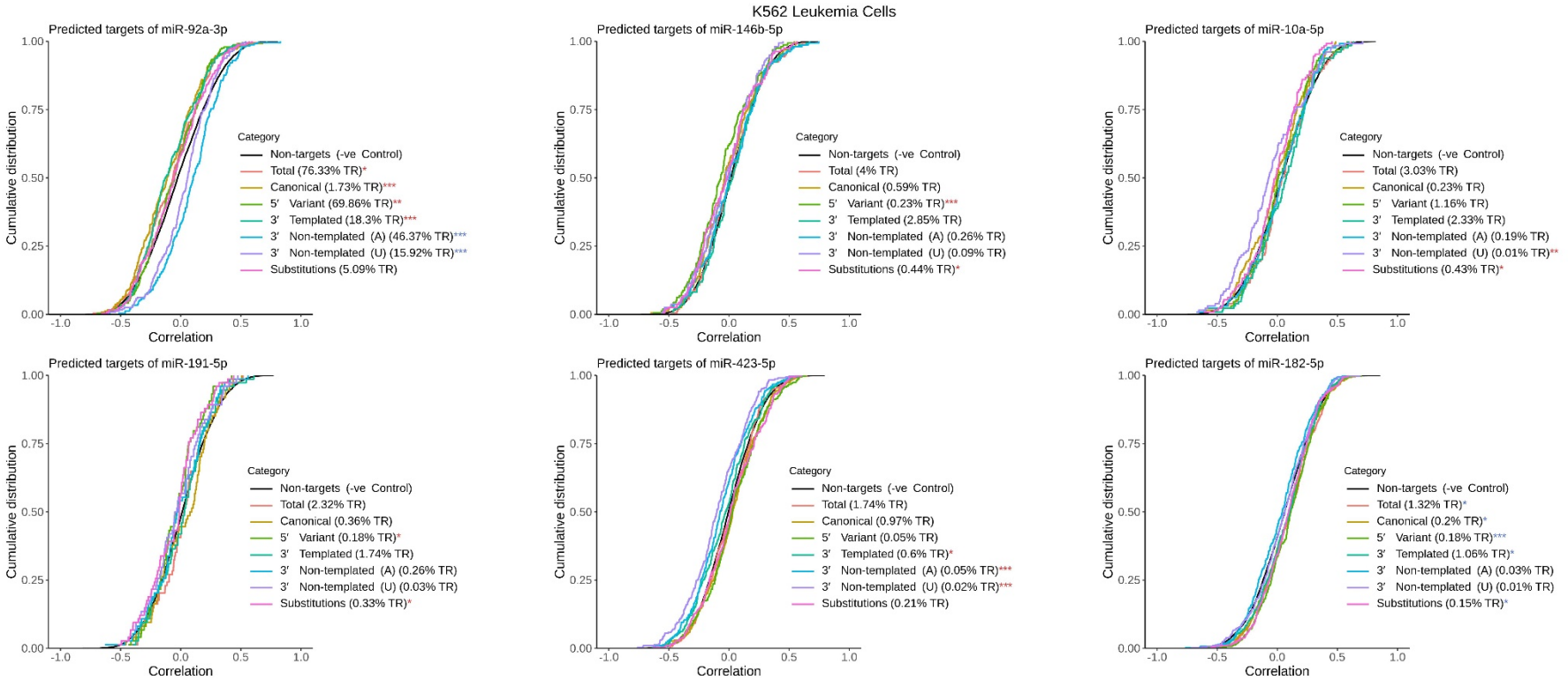

Figure 3.9. Correlation of expression levels between canonical miRNAs or isomiR categories and the canonical miRNAs predicted targets (mRNAs) in K562 Leukemia cells. In the same miRNA, isomiR categories vary significantly in correlation strength and direction, suggesting these modifications lead to differences in regulatory activity. The 6 highest expressed miRNAs are shown excluding those with low isomiR abundance. For each miRNA, an aggregated list of predicted target mRNAs was collected using miRNAtap and expression of targets was correlated against normalized counts for each isomiR category. Percent of total reads for each isomiR category are shown (averaged across all K562 cells). Non-targets included all remaining mRNAs which were not predicted as a target for that miRNA. Significant differences between each category and Non-targets were calculated using the two-sided Kolmogorov-Smirnov test (p-value < 0.05) and are indicated with a blue (positive correlation) or red (negative correlation) asterisk. **TR:** Total Reads (mapped to miRNAs).

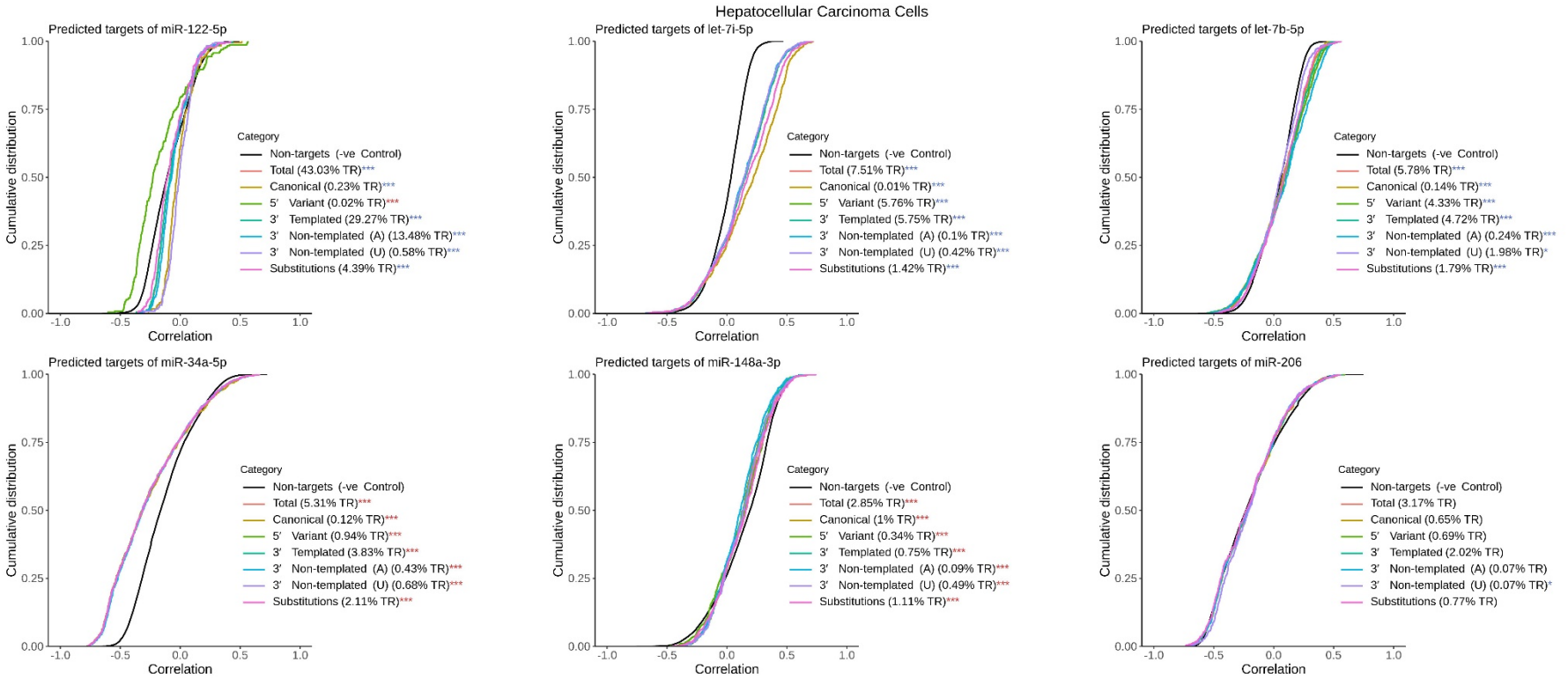

Figure 3.10. Correlation of expression levels between canonical miRNAs or isomiR categories and the canonical miRNAs predicted targets (mRNAs) in Hepatocellular Carcinoma (HCC) cells. In the same miRNA, isomiR categories vary significantly in correlation strength and direction, suggesting these modifications lead to differences in regulatory activity. The 6 highest expressed miRNAs are shown excluding those with low isomiR abundance. For each miRNA, an aggregated list of predicted target mRNAs was collected using miRNAtap and expression of targets was correlated against normalized counts for each isomiR category. Percent of total reads for each isomiR category are shown (averaged across all HCC cells). Non-targets included all remaining mRNAs which were not predicted as a target for that miRNA. Significant differences between each category and Non-targets were calculated using the two-sided Kolmogorov-Smirnov test (p-value < 0.05) and are indicated with a blue (positive correlation) or red (negative correlation) asterisk. **TR:** Total Reads (mapped to miRNAs).

3.6. Unique Molecular Identifiers increase estimated isomiR abundance

A key difference in the Small-seq protocol compared to the Holo-seq and Co-seq protocols was the incorporation of unique molecular identifiers (UMIs). In bulk studies, UMIs have been found to reduce cDNA amplification bias during library preparation and enable a more accurate quantification of the transcriptome[261,277]. However, to our knowledge their effect on isomiR abundance has not been thoroughly investigated. Therefore, we assessed how UMI deduplication might influence isomiR expression in the glioblastoma cell lines, HEK293 and embryonic stem cells. We found that in all cell types, the relative abundance of canonical miRNAs was decreased after UMI deduplication (Figure 3.11). The corresponding effect on all isomiR categories typically trended towards an increase in relative expression levels, with 3p template extensions (+) in naïve primed embryonic stem cells being the only significant exception (Figure 3.11). While consistent, the overall effect after UMI deduplication was modest for most cell types and isomiR categories, however there were notable exceptions such as the naïve embryonic stem cells which had more than a 2-fold increase in extended 5’ variants as well as a notable increase in shortened 5’ variants and 3´ template variants. Correspondingly, there was a clear reduction in extended 3´ template variants.

In most cases, UMI deduplication had a minor impact on the overall shape of the peak distributions (Figure 3.12 and 3.13). However, in the primed embryonic stem cells there was a clear shift in the peak length from 23nt to 22nt after deduplication which led it to resemble the other Small-seq cell types more closely. When comparing the effect of UMI deduplication on miRNA lengths across each isomiR category, we found the increase in measured isomiR expression was not equal across all categories. 5p variants and substitutions showed more obvious increases in expression, and the expression of shorter isomiRs were generally affected to a higher degree after deduplication. We also compared the averaged proportions of positional modifications between non-deduplicated and deduplicated cell types but did not observe any differences (Figure 3.14). Overall, these findings suggest protocols which do not utilize UMIs may underestimate the expression of most isomiRs.

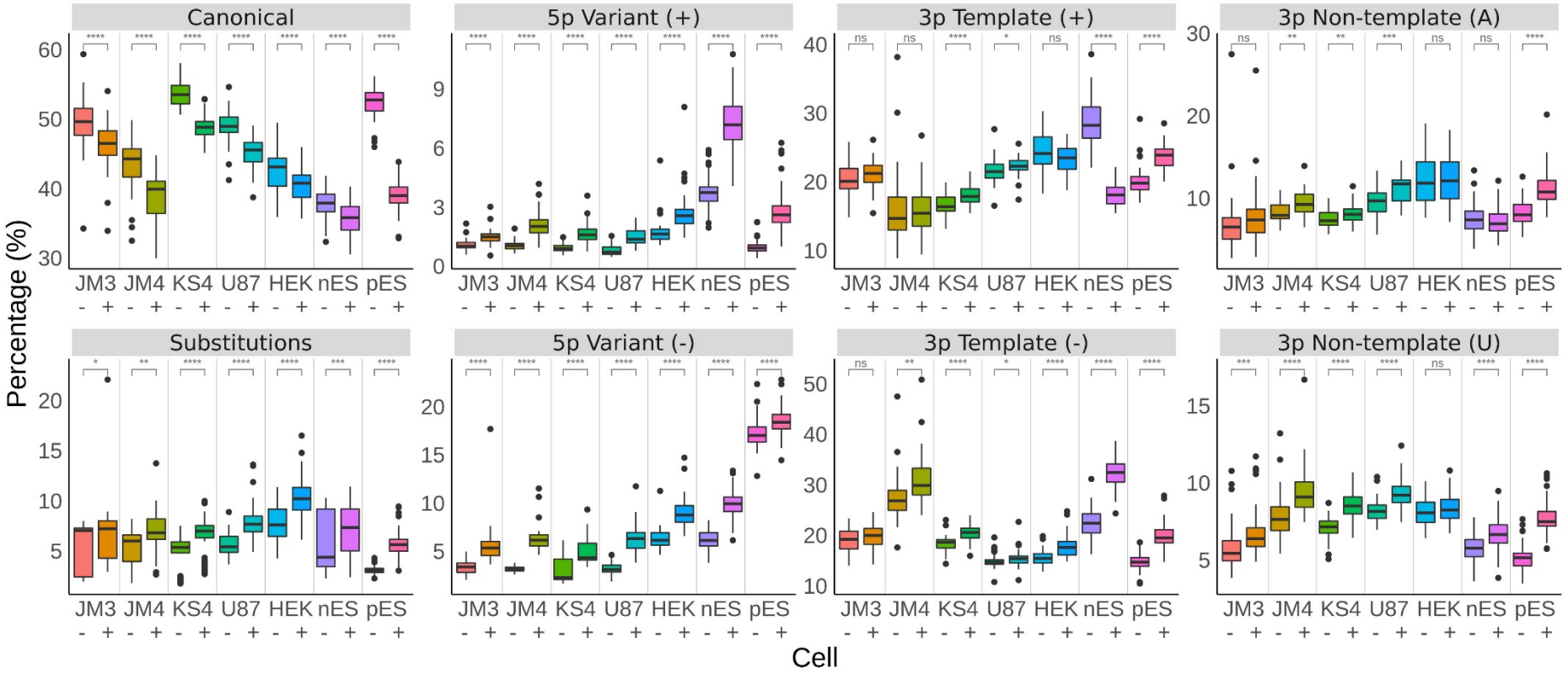

Figure 3.11. Impact of UMI deduplication on isomiR types. For each cell non-deduplicated reads (-) are shown next to deduplicated reads (+). **ns:** Not significant. **nES:** Naïve embryonic stem cells. **pES:** Primed embryonic stem cells.

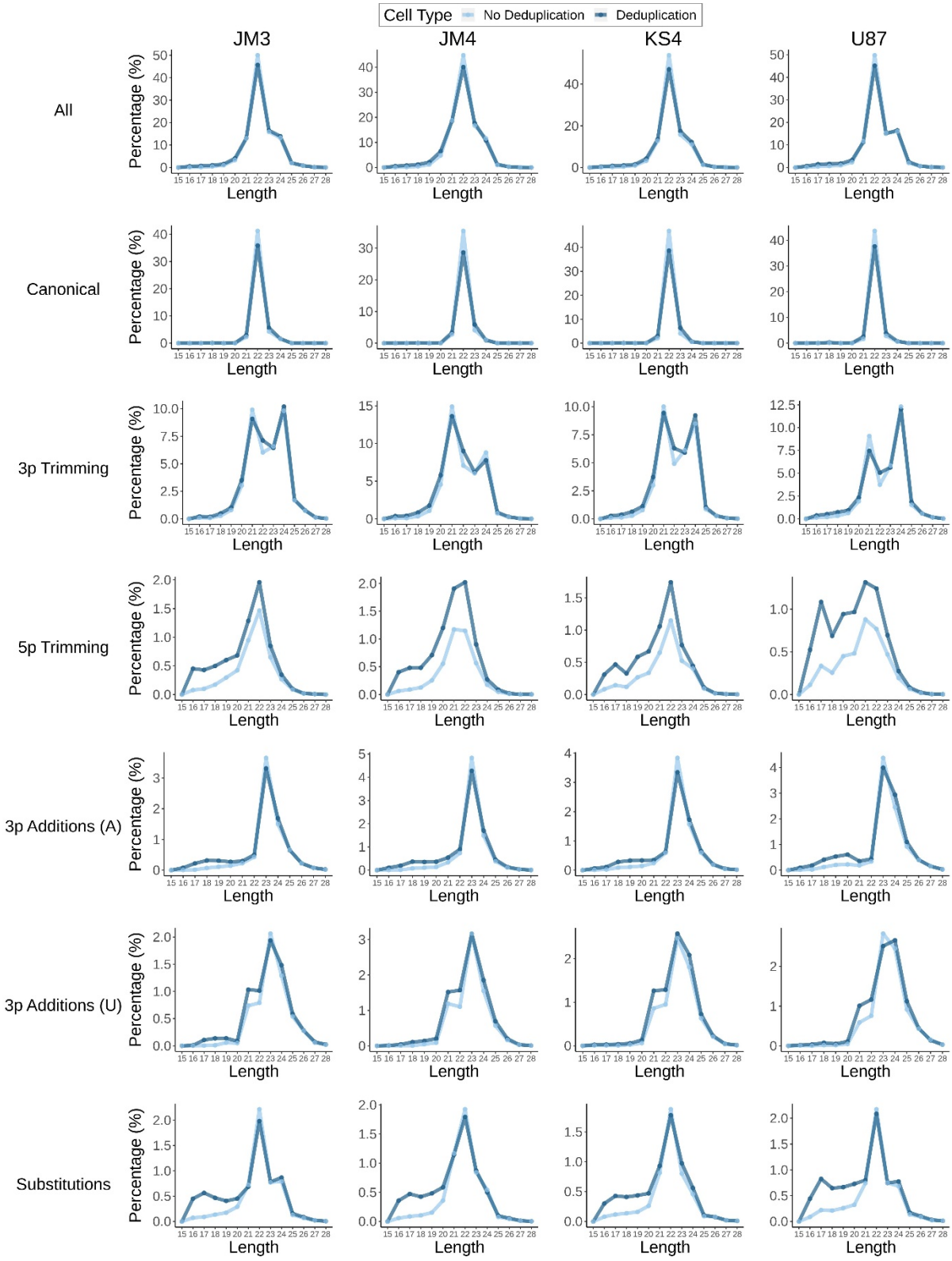

Figure 3.12. Impact of UMI deduplication on miRNA length distributions in glioblastoma cell lines. Shows read lengths without deduplication (light blue) and with UMI deduplication (dark blue).

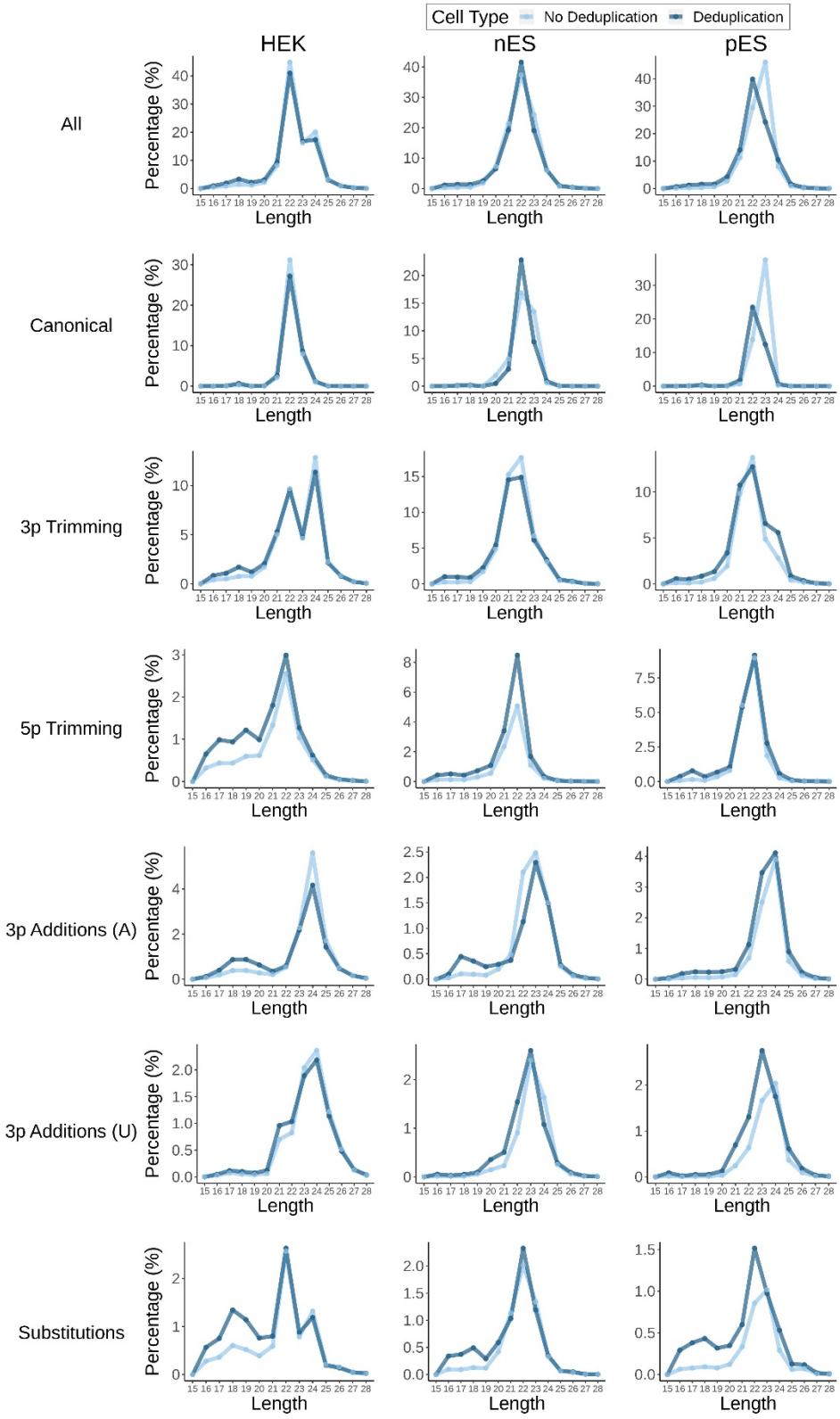

Figure 3.13. Impact of UMI deduplication on miRNA length distributions in the HEK293 cell line, and naïve and primed embryonic stem cells. Shows read lengths without deduplication (light blue) and with UMI deduplication (dark blue).

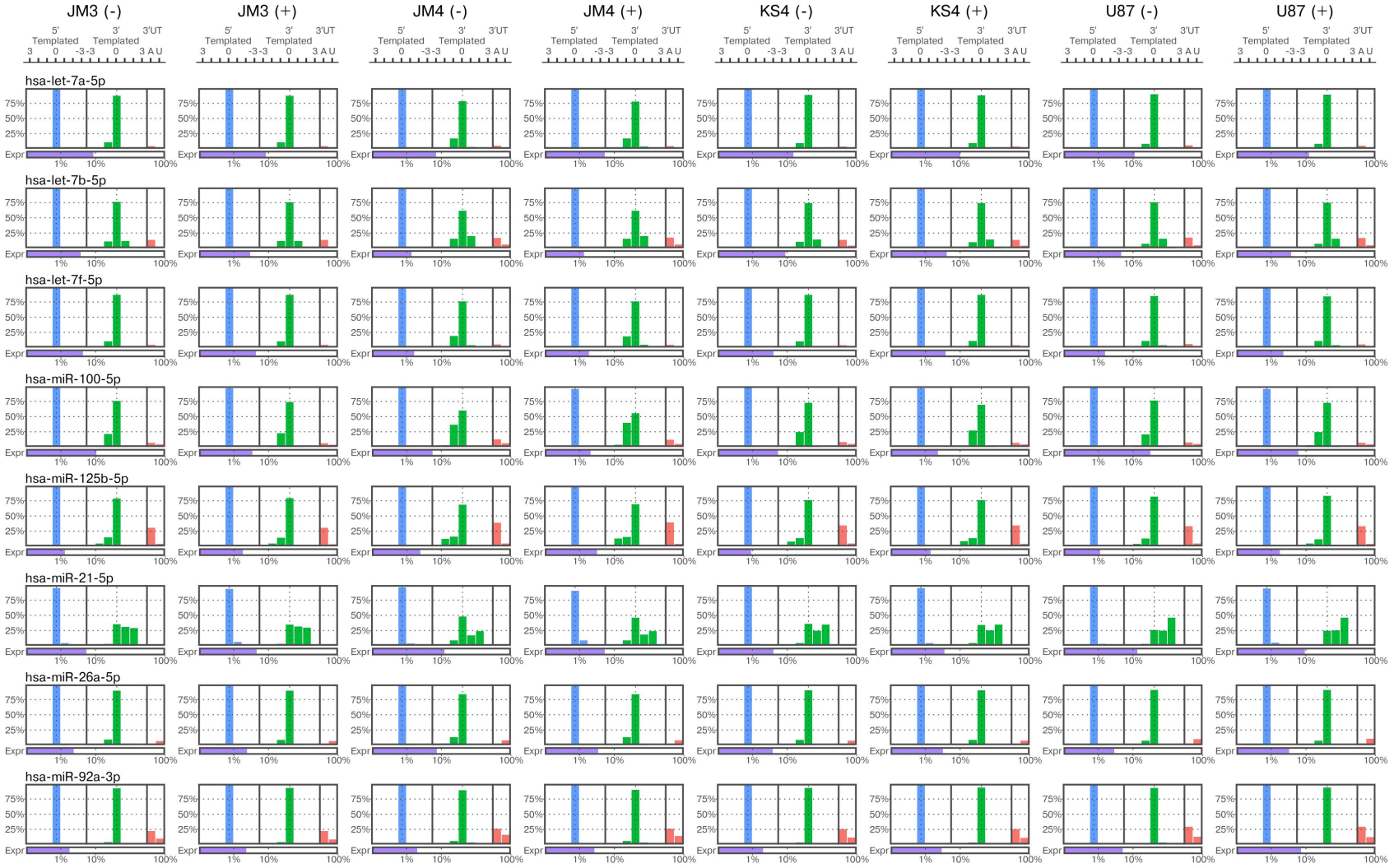

Figure 3.14. Comparison of 5′ and 3′ modifications between averaged non-deduplicated (-) and deduplicated (+) glioblastoma cells, for selected miRNAs.

3.7. Discussion

In this study we highlighted major differences in the way isomiRs are expressed or processed in various cell types, supporting the cell autonomous regulation of isomiRs. When comparing the processing of different miRNAs, the single cell data in this analysis mirrored observations with bulk sequencing data from other studies which showed sequence modifications vary substantially when comparing different miRNAs. Evidently, properties specific to each miRNA dictate how it interacts with the cell's miRNA or isomiR processing machinery, and may be influenced by nucleotide sequence, secondary structure, interactions with Argonaute or the miRNA's targets[62,278,279]. Collectively, these findings highlight isomiR processing as a novel gene regulatory layer whose function has not been fully appreciated with population level isomiR studies. We argue that future studies investigating isomiR expression and function should consider incorporating single cell methods to better understand their role. However, the lack of technological development in single cell small RNA sequencing remains a major constraint and cheaper, high throughput methods are likely necessary to support more comprehensive studies.

3.7.1. Analysis of isomiR types overcomes current limitations with single cell small RNA sequencing data

There are several challenges that we had to consider when analysing isomiR expression in single cell sequencing data. This included the low number of reads which mapped to miRNAs for each cell, as a high proportion of reads tend to map to ribosomal RNA or do not even map to the human genome, possibly due to low level contamination. Low target read counts are a typical problem in single cell sequencing studies, however miRNAs in small RNA-seq often represent an even smaller proportion of total reads than messenger RNAs in RNA-seq[168,188]. This issue can be further compounded with isomiR studies which separate a miRNA's reads into each unique sequence, creating a large number of isomiRs with extremely low counts. Additionally, the presence of errors from prior stages such as library amplification and sequencing can lead to spurious isomiRs which overestimate the number of real isomiRs[280,281].

Our approach to address these issues was to focus on the type of isomiR modifications at the 5´ or 3´ ends and quantify isomiRs based on biologically relevant isomiR types. There are at least two benefits to this approach. Firstly, reads are no longer split across a large number of features and each isomiR type will generally have more reads contributing to them, therefore are less influenced by variability and drop-outs[282]. Secondly, this meant that expression of each isomiR type would only be confounded by errors in a smaller region (e.g 5´ or 3´ end) and not by errors across the remainder of the sequence. We were able to show a relationship between expression of various isomiR types and their miRNA's canonical targets in highly expressed miRNAs, implying this single cell small RNA sequencing data can

be used to predict isomiR function when applying this methodology. A limitation of this approach is that it may not capture the full range of biological functions driven by isomiRs, for example if some isomiRs are only functional when certain combinations of isomiR modifications are present, or if their function depends on sequence specific interactions with their targets and/or Argonaute. Nonetheless, this may be a more suitable approach for studying isomiRs in single cell data until a greater depth of miRNA reads becomes feasible and spurious isomiRs can also be accounted for.

3.7.2. Biases in small RNA sequencing protocols are likely to influence isomiR quantification

Since both the Co-seq and Holo-seq studies only sequenced one human cell type each, and there were no shared cell types across studies, we were unable to ascertain how much of these differences could be attributed to biological or experimental factors. However, the high amount of isomiRs in the single cell data for HCC and K562 cells was not reflected in their respective bulk sequencing data obtained from several independent studies (Figure 3.3) and were unable to find further evidence that this was a feature of these cell types. However, it is possible that this isomiR expression profile is characteristic of a smaller cell population that was not evident with bulk or population level sequencing.

Previous comparisons of small RNA sequencing protocols have demonstrated substantial biases depending on the library preparation method. For example, the Clontech small RNA sequencing kit utilizes an adapter-free ligation method and was shown to detect a high number of false isomiRs when compared to the adapter ligation-based NEXTflex kits[283]. Many small RNA sequencing protocols, including the three single cell methods included in this study, utilize an adapter ligation-based library preparation[188,191,194,283–285]. This approach begins with a 3´ ligation step involving a pre-adenylated DNA oligonucleotide, followed by 5´ ligation of an RNA oligonucleotide for small RNA capture. The RNA/DNA-joined molecules are then reverse transcribed and amplified with PCR. However, the terminal nucleotide sequences of adapters are known to influence their ligation efficiency with small RNAs at both the 5´ and 3´ ends due to a change in RNA structure[286]. Given that the adapter sequences including the terminal bases were different in each of the protocols, this may be one factor that influenced the detection and quantification of isomiRs. There were also key differences in the sample preparation and RNA isolation procedures for each protocol which may have also influenced isomiR detection and quantification. Although the global expression profiles of cell types from the small-seq protocol were similar, there were still significant differences in isomiR type expression when analysing individual miRNAs which indicates that the sequencing protocol cannot fully explain the observed differences. This suggests that both cell type and protocol can influence isomiR expression measurements. More comprehensive studies will be needed to understand how widespread isomiR variation is across cell types in healthy and diseased tissues.

3.7.3. Methods to improve single cell small RNA sequencing data quality for isomiR analysis

Even if present at low frequencies, sequencing artefacts are a particular concern in isomiR studies, as single base alterations can completely redefine an isomiR's type and interpreted function. Implementing UMI's into the sequencing library can assist with more accurate measurements of expression levels and enable more confidence in isomiR classification[261,287]. Additionally, ligation biases can be reduced by incorporating random bases at the ends of the adapters which are ligated to RNAs, or on splints which can be introduced to facilitate ligation between adapters and small RNAs[288,289].

3.7.4. Implications of cell autonomous isomiR regulation in cancer

Previous studies have already investigated the potential for isomiRs to classify cancers with a high degree of success[142,200]. As this research has traditionally relied on bulk sequencing data to capture population level isomiR expression, it is not clear if this inter-tumour heterogeneity extends to intra-tumorigenic differences between cancer cells. Our findings indicate cells possess autonomous isomiR regulation and support a broader function in gene regulation. Furthermore, we found that individual cells from the same cell type were relatively homogenous when considering cell lines and primary cultures, compared to the more heterogeneous isomiR expression was observed in the resected HCC tumour cells. Although there is insufficient data to draw conclusions, this could be an indication that cells existing in their native environment may be influenced by a combination of cell autonomous and cell-extrinsic factors which tune isomiR expression and contribute to inter-cellular variation. Collectively, our work highlights that isomiRs are functional in single cells and their regulation is cell autonomous, which necessitates further research as this may be an unexplored contributor to intra-tumoural heterogeneity.

4. Investigating the RNA binding properties of alternatively spliced Argonaute 2 (Ago2) isoforms

4.1. Chapter introduction

MicroRNA (miRNA) biogenesis involves various proteins to process primary miRNA transcripts into mature miRNAs that can regulate gene expression[1]. The Microprocessor complex formed by Drosha and DGCR8 cleaves primary miRNAs into shorter double stranded precursor miRNAs with a stem-loop structure[5]. Precursor miRNAs are exported into the cytoplasm and recognized by Dicer, which removes the loop to release a miRNA duplex[1]. Argonaute then binds the miRNA duplex and selects one strand for incorporation (guide strand) into the RNA induced silencing complex (RISC) while the other is ejected and degraded (passenger strand)[67].

Argonaute is an essential component of the RNA induced silencing complex (RISC) and is required for miRNA-mediated gene regulation[11,67,68]. While the sequence of the miRNA is critical for target recognition, Argonaute provides miRNAs with their function by directly cleaving targets or recruiting various proteins to suppress gene expression through translational inhibition and deadenylation[278,290]. Argonaute also affords miRNAs protection from cytoplasmic nucleases and speeds up target searching by changing the conformation of the miRNA[279,291]. In humans, Argonaute proteins are expressed from four different genes (AGO1, AGO2, AGO3 and AGO4)[292]. Although there are minor differences in sequence, each Argonaute contains identical protein domains and appear mostly redundant in terms of RNA binding and function, except for Ago2's slicer activity[292].

4.1.1. Argonaute loads multiple classes of small RNAs

Although miRNAs typically form the bulk of small RNAs that are loaded into Argonaute, many other classes of small RNAs can form a functional RISC through non-canonical biogenesis pathways[293]. This includes ribosomal RNA-derived fragments, transfer RNA-derived fragments, small nuclear RNAs, small nucleolar RNAs and other classes of small RNAs[293].

Functionally distinct Argonautes that bind to different classes of small RNAs have been described in the fruit fly Drosophila melanogaster and nematode Caenorhabditis elegans. Drosophila expresses two structurally and functionally distinct Argonautes, Argonaute 1 (AGO1) and Argonaute 2 (AGO2). AGO1 binds and mediates miRNA-directed cleavage whereas AGO2 binds siRNAs and is essential for siRNA-directed cleavage[294]. In C. elegans there are 19 different functional Argonautes which either participate in miRNA-mediated gene regulation or bind other classes of small RNAs to function as part of a broader genomic immune system that suppresses foreign RNA or endogenous retroviruses and transposons[295]. Although studies have found that the four Argonaute genes in humans tend to bind

similar profiles of small RNA classes and miRNAs, they typically rely on antibodies which are unable to distinguish between their isoforms[67,296]. Many Argonaute isoforms have been annotated in the reference sequence databases RefSeq (National Center for Biotechnology Information) and GENCODE (GENCODE consortium)[297,298]. This includes isoforms with variations in their protein coding sequence that still possess the functional domains essential for RISC formation and gene regulation[67]. Therefore, it is possible that multiple Argonaute isoforms function in gene regulation and may have distinct small RNA binding characteristics which alter their regulatory activity. However, to date the biological significance of any Argonaute isoforms have not been investigated.

4.1.2. Features of the miRNA duplex influence Argonaute strand selection

Nearly all miRNA genes encode two miRNAs with distinct nucleotide sequences, enabling alternate sets of regulatory targets depending on which miRNA will be incorporated into the RISC[299,300]. miRNA strand selection by Argonaute is not random and is known to be influenced by sequence and structural features pertaining to the miRNA duplex[299]. Primarily, two properties of the miRNA duplex contribute to strand selection – the relative thermodynamic stability of each end and the 5´-terminal nucleotides of each strand.

Early studies with siRNAs and miRNAs found that strand selection was dependent on the relative thermodynamic stability of each end of the precursor duplex. Argonaute prefers to unwind duplexes on the side with weaker pairing, resulting in the incorporation of the strand whose 5´ end has been unwound[299]. The importance of thermodynamic stability was first described by Schwartz et al who found that altering the relative thermodynamic stability of each end through sequence mismatches could change which strand was incorporated into Argonaute[301]. The impact of sequence mismatches was limited to the first 4 paired nucleotides of each end of the miRNA duplex. However, later studies found that thermodynamically unfavourable strands were often highly expressed and sometimes more dominant in tissues, suggesting other factors also contributed to strand selection[302].

Subsequent studies found Argonaute has a strong bias towards certain bases at the 5´-terminal end of each miRNA strand. Hu et al highlighted a strong bias favouring uridine bases in the first nucleotide of 33 expressed miRNA pairs[303]. Conversely, they found cytosines were strongly unfavored in the first nucleotide[303]. Frank et al found a strong bias for uridine or adenine and against cytosine or guanine[304]. This bias was attributed to a rigid loop structure housed in the MID domain of Argonaute, which interacts with the phosphorylated 5´ end of the miRNA guide strand during loading[304]. To date, no study has investigated if alternative splicing can change strand selection bias of Argonaute to determine which miRNAs participate in gene regulation.

4.1.3. Alternative splicing may produce Argonaute proteins with altered functions

In this study we investigate alternatively spliced isoforms of Argonaute 2 in humans and highlight 3 isoforms which are expressed in human tissues and possess distinct RNA binding properties. This includes a reduced affinity towards most miRNAs in two isoforms, compared to the canonical Argonaute 2 which is studied most extensively, suggesting that alternative splicing may reduce miRNA-mediated regulation in some tissues. Additionally, we identified several miRNAs with increased abundance, relative to other miRNAs, within non-canonical Ago2 isoforms and find evidence that this may be caused by a reduction in 5´ nucleotide bias that alters miRNA strand selection in some cases. Our work identifies alternative splicing as a novel mechanism which can regulate gene expression by tuning miRNA selection. This further highlights the complexity of the miRNA pathway and has implications in gene regulation in normal and cancerous cells.

4.2. Specific methodology

4.2.1. Buffers

- PBS (Phosphate Buffered Saline) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4)
- PBST (Phosphate Buffered Saline Tween) (1X PBS + 0.1% Tween 20)
- NP-40 lysis buffer (20mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5% NP-40, 2mM EDTA, 0.5 mM DTT, 1 mM NaF, 1 Protease Inhibitor Tablet)
- IP Washing Buffer (50mM Tris-HCl pH 7.5, 300nM NaCl, 5mM $MgCl_2$, 0.05% NP-40)
- PK Buffer (200mM Tris-HCl pH 7.5, 25mM, 300 mM NaCl, 2% SDS w/v)

4.2.2. Reagents

- DMEM (Dulbecco's Modified Eagle Medium) (4.5 g/L D-Glucose, L-Glutamine, 25mM HEPES) (Gibco)
- Opti-MEM + GlutaMAX-l (HEPES, 3.024 g/L Sodium Bicarbonate) (Gibco)
- Trypsin TrypLE™ Select Enzyme (Gibco)
- DPBS (Dulbecco's Phosphate Buffered Saline (Gibco)
- FBS (Fetal Bovine Serum) (Gibco)
- Penicillin-Streptomycin (Invitrogen)
- Lipofectamine 2000 (Invitrogen)
- Trypan Blue Solution (Gibco)
- Dynabeads Protein G (Invitrogen)
- Phenol-Chloroform-Isoamyl Alcohol 25:24:1 (Sigma)
- Proteinase K (Invitrogen)
- Protease Inhibitor Tablets (Thermo Scientific)

4.2.3. Antibodies

Target	Product	Dilution
Ago2	Rat Monoclonal Anti-Ago2 clone 11A9 (Sigma Aldrich)	1:1000
FLAG	Mouse Monoclonal ANTI-FLAG® M2 (Sigma-Aldrich)	1:1000
Mouse IgG	Donkey Polyclonal Secondary Antibody to Mouse IgG (Abcam)	1:5000

Table 4.1. Antibodies used for immunoprecipitation experiments

4.2.4. Cell provision

HeLa and HEK293 cells were sourced from the American Type Culture Collection (ATCC). THP-1 cells were obtained from Bernadette Sanders from the Faculty of Science at UTS. SCC38, SCR4, LNCaP, and PNT2 cell lines were obtained from Nham Tran from the School of Biomedical Engineering at UTS.

4.2.5. Ago2 isoform plasmid transfection

All wet lab experiments were performed by Sarah Bajan. HeLa cells were grown to confluency in 15 cm culture dishes for the transfection. Before transfection, old media was removed, cells were washed with DPBS, and then fresh antibiotic-free DMEM.

All Ago2 isoform plasmids were synthesized and cloned into N-terminal FLAG-tagged CMV 10 plasmids through Sigma-Aldrich's gene synthesis service. A transfection mix was prepared with each of the Ago2 isoform plasmids. Transfection mix was prepared by creating a DNA plasmid solution diluted with Opti-MEM and a 0.8% Lipofectamine 2000 solution diluted in Opti-MEM, each at volume of 3.5 ml per confluent 15cm dish. 10 ng of plasmid was used for Ago2 Isoform 1, 15 ng for Isoform 2, and 30 ng for Isoform 3. Concentrations were adjusted to produce similar concentrations of Ago2 protein as determined by Western blot. The DNA plasmid and lipofectamine mixtures were incubated separately for 5 mins, then mixed and incubated for an additional 15 min. 7 ml of the transfection mix was distributed across each dish and lightly swirled, then incubated for 5 hours. Afterwards, the transfection media was then removed, cells washed with DPBS, and fresh antibiotic-free media was added. Cells were incubated for 24 hours after transfection before being harvested.

4.2.6. Cell lysis for immunoprecipitation

For transfected cells prepared for immunoprecipitation, cells were harvested by adding 2 ml of NP-40 lysis buffer and using a cell scraper to dislodge cells from the dish. The cell lysate was then moved to Eppendorf tubes and incubated at 4°C for 15 min while rotating. Lysis was confirmed with Trypan Blue solution. After incubation, the lysates were centrifuged at 14000g for 5 min, then the supernatant was moved to a new tube. For each sample, 40 µl was removed for Western blot to confirm successful transfection and the remaining sample was snap frozen with liquid nitrogen and stored at -80°C until immunoprecipitation.

4.2.7. Immunoprecipitation of FLAG-tagged Ago2 isoforms

Prior to immunoprecipitation, 100 µl per sample of Dynabeads Protein G were washed three times with 1ml PBST for 1 min using a magnetic rack to remove supernatant. The Dynabeads Protein G was then split in half, with one half mixed with 10 µg of either FLAG or 11A9 antibody and the remaining 50 µl mixed with 10 µg of IgG antibody as a control. The antibody-bead mixtures were incubated at 4°C for 30 min. Beads were then washed three times with 1ml PBST for 2 min using a magnetic rack

and then resuspending in 50 µl PBST. 50 µl of antibody-bead mixture was added to each lysate for immunoprecipitation and incubated at 4°C overnight. Afterwards, the supernatant from each sample was isolated for Western blot and beads were washed three times in 1ml IP Washing Buffer for 2 min using a magnetic rack. After washing, the immunoprecipitated beads were resuspended in 100 µl IP Washing Buffer. 90 µl of the immunoprecipitated beads were used for Proteinase K digestion and RNA extraction while the remaining 10 µl was reserved for Western blot. Proteinase K digestion was performed by adding 210 µl of PK Buffer, 40 µg Proteinase K, and 1 µl Glycogen to each sample and incubating at 65°C for 20 min.

4.2.8. RNA extraction

For RNA extraction, 300 µl of phenol-chloroform was added to each sample, vortexed, then centrifuged at 16000g for 15 min at 4°C. The aqueous layer containing RNA was removed and added to 900 µl of 100% ethanol. Samples were left overnight at -80°C for precipitation. The next day, samples were centrifuged at 16000g for 30 min at 4°C. The supernatant was removed, and the remaining RNA pellet was washed with 1 ml of 75% ethanol. Samples were centrifuged at 16000g for 15 min at 4°C, then the supernatant was removed, and the RNA was air dried for approximately 5 min or until it became translucent. RNA was then resuspended in 20-50 µl of RNase-free water and the RNA concentration was measured using the Nanodrop according to manufacturer's instructions. Samples were snap-frozen using liquid nitrogen and stored at -80°C.

4.2.9. Sequencing data generation

Sequencing data for Ago2 isoform quantification was generated from a previous Capture-seq experiment by the Hutvagner lab. RNA from cell lines and tissues were sent to the Ramaciotti Centre for Genomics for Capture-Seq with a gene panel that included Ago2. RNA for the human tissues was sourced from Clontech as part of the Human Total RNA Master Panel II, except for Heart tissue which was bought separately.

For generating the Ago2 isoform small RNA sequencing data with HeLa cells, wet lab experiments were performed by Sarah Bajan which included 2 biological replicates for each isoform.

4.2.10. Calculation of Percent Spliced with Junction (PSJ) values

To estimate the proportion of Ago2 transcripts containing an exon junction, a STAR index was prepared with RefSeq's human genome (GRCh38) and with default parameters except: --sjdbOverhang 150[305]. Raw reads from the Capture-seq dataset were then aligned with STAR to this index with default parameters. The spliced junction file (SJ.out.tab) from STAR's output was used to quantify the Percent Spliced with Junction (PSJ) values, determined by considering the unique read

count at each relevant exon junction, divided by the total reads belonging to junctions overlapping the selected exon junction's genomic region.

4.2.11. Small RNA biotype quantification

Raw reads were trimmed using cutadapt (v2.7) and reads shorter than 10 nucleotides or containing any N bases were removed. Reads were then mapped sequentially using bowtie (v1.2.3), with the following parameters: n 2 -e 120 -l 20 --best[228,229]. Reads were first mapped to miRNAs, then the remaining unmapped reads were mapped to ribosomal RNAs (rRNAs), followed by transfer RNAs (tRNAs), then remaining small RNAs. Bowtie indexes for miRNAs were generated using miRBase's precursor hairpins (v22.1), for rRNAs using merged sequences from filtered rRNAs in Gencode's primary annotations (v35) and Arb-silva's LSU and SSU rRNA databases (v138.1), for tRNAs using filtered tRNAs in Gencode's primary annotations (v35), for remaining small RNAs using filtered small RNA biotypes in Gencode's primary annotations (v35) – including miRNA, misc_RNA, rRNA, scRNA, snRNA, snoRNA, ribozyme, sRNA, scaRNA, and vaultRNA[36,298,306]. The resulting sam files were merged and reads mapping to each biotype were quantified.

4.2.12. miRNA and isomiR quantification

Raw reads were trimmed using cutadapt (v2.7) and reads shorter than 10 nucleotides were removed[228]. Data was converted from FASTQ to FASTA file format and reads were collapsed so that each unique sequence only appeared once in each file and counts were appended to the headers. Collapsed reads were aligned to miRbase (v22.1) annotated precursor miRNAs using miraligner (v3.4), with the following parameters: -sub 1 -trim 3 -add 3[36,231].

For quantification of isomiR types, the following categories were defined: Canonical – miRNAs with a perfect match to the miRbase mature sequence, 5′ Variant – miRNAs differing at the 5′ end with respect to the miRbase mature sequence, 3′ Template – miRNAs deviating at the 3′ end to the miRbase mature sequence but still matching the precursor miRNA sequence, 3′ Non-template – miRNAs which did not match the precursor miRNA sequence at the 3′ end, and Substitution – miRNAs containing a maximum of 1 mismatch to the miRbase mature sequence, excluding variations at the 5′ or 3′ ends. Categories were then assigned any alignments which contained their respective isomiR variation and were used for measuring their proportions relative to the total number of annotated reads.

4.2.13. Differential expression analysis

Differential miRNA expression between the Ago2 isoforms was determined using the quantified sequencing data from the Ago2 isoform immunoprecipitations and the DESeq2 R package (v1.26.0), filtering results by baseMean >= 1000, $\log_2$ fold-change >= 3, and adjusted p-value < 0.05[307].

4.2.14. Estimating base preferences in Ago2 isoforms

Strand selection probabilities across all miRNAs were estimated by comparing the first base (nucleotide) of each mapped miRNA, considered a guide strand, to its predicted passenger strand, and for each pair of bases calculating the proportion of events where one base belonged the guide and the other to the passenger strand. The miRbase.db R package (v1.2.0) was used to get the nucleotide sequences for all passenger strands[36]. 5´ isomiRs were excluded as we could not determine if the isomiR was produced before or after strand selection. For calculations for individual miRNAs, miRNAs belonging to precursors without two annotated strands were not included in the analysis. Probabilities of any pairs of bases calculated from less than 20 mapped reads were also excluded.

4.2.15. Code for Data Analysis and Figures

Documents containing code used to generate the results in this section can be found in the following link: https://cloudstor.aarnet.edu.au/plus/s/GaNTw7rMWztjliu

4.3. Alternative Splicing of Ago2 Suggests Diversity of Isoform Expression in Human Tissues

To date, expression of Argonaute 2 (Ago2) isoforms have been poorly characterized in human tissues and cells, and it is unknown which transcript variants are biologically significant (Figure 4.1A). In a previous experiment from our lab, a targeted panel of genes including Ago2 were enriched and sequenced using Capture-Seq[308]. As this data spanned 16 different human tissues and cell lines, we were able to estimate the relative frequency of alternative splicing events that characterised the Refseq annotated transcripts of Ago2 (Figure 4.1B; Appendix C)[297]. We used data from reads overlapping each exon junction to calculate a Percent Spliced with Junction (PSJ) value, estimating the proportion of Ago2 transcripts containing each splicing event. As our analysis identified reads supporting several unannotated splice junctions, we also calculated PSJ values for the top 5 unannotated junctions by read count across all samples.

A wide range of PSJ values were observed between samples for all exon junctions, revealing extensive splicing across different biological contexts (Figure 4.1B). Our data indicated that the most extensive splicing was at the 5´-end of Ago2. Three of the Refseq annotated transcripts (XM_011516965, XM_011516966, and XM_011516968) contain alternative starting exons compared to the wild-type transcript NM_012154 (Figure 4.1A). PSJ values (Figure 4.1B) for the wild-type (NM_012154) exon junction between exon 1 and 2 (red) indicated that, in all samples, splicing of this junction was the most dominant across this region, however there was considerable variation across samples where it was lowest in the LNCaP cells (59.86%) and highest in the skeletal muscle (97.29%). Our analysis supported the presence of two of the annotated transcripts, XM_011516965 (Chr8:140585312-140596675; yellow) and XM_011516966 (Chr8:140585312-140605796; cyan), in multiple cell lines and tissues, as well as at least three unannotated transcripts (grey) due to the presence of novel exon junctions with exon 2 of the wild-type transcript (Chr8:140585312-140588721, Chr8:140585312-140590593 and Chr8:140585312-140633774; grey). There was a paucity of reads spanning the regions near each unannotated junction and any other exon, suggesting these splicing events could also be alternative starting exons, and not exon inclusion events between exon 1 and 2. Among these exon junctions, representing putative alternative starting exons (excluding the wild-type junction), each had the highest PSJ value in at least one sample, suggesting cell or tissue specific functions for these splicing events.

The Refseq annotations also describe two transcripts which are characterised by exon skipping events at exon 2 (XM_017013317) and exon 17 (NM_001164623) in relation to the wild-type transcript. Reads spanning the exon junctions that represent these skipping events were identified in nearly all samples.

For the exon junction Chr8:140572933-140635484, uniquely associated with XM_017013317 (blue), PSJ ranged from 0% in whole brain and testes up to 5.77% in kidney. Splicing of the exon junction Chr8:140532616-140539319, uniquely associated with NM_001164623 (green), was relatively low with PSJ ranging from 0% in foetal brain, bone marrow, and testes, to 0.85% in LNCaP cells.

Finally, we highlighted two additional unannotated exons junctions which were evidence of an alternative donor splice junction site and inclusion of a novel exon. There was an alternative donor (5´) splice site identified between exon 6 and 7 (Chr8:140558573-140559419), which was almost exclusively expressed in bone marrow tissue with a PSJ of 4.23%.

A novel exon junction was identified connecting to exon 3 (Chr8:140572604-140572811), which ranged in PSJ from 0% in foetal brain, whole brain, and testes tissue to 9.36% in bone marrow. We did not find evidence of reads connecting this novel exon with downstream exons, suggesting this may lead to a truncated transcript.

Collectively, these results suggest that transcripts from the Ago2 gene are extensively spliced and that many of these splicing events are altered in various biological contexts. Following this, our study focused on three Ago2 transcripts - NM_01254.5 (Isoform 1), NM_001164623.3 (Isoform 2), XM_017011317.2 (Isoform 3), which are likely to produce proteins with altered functional domains (Figure 4.1C) and investigated functional differences through their association with small RNAs (Figure 4.1B).

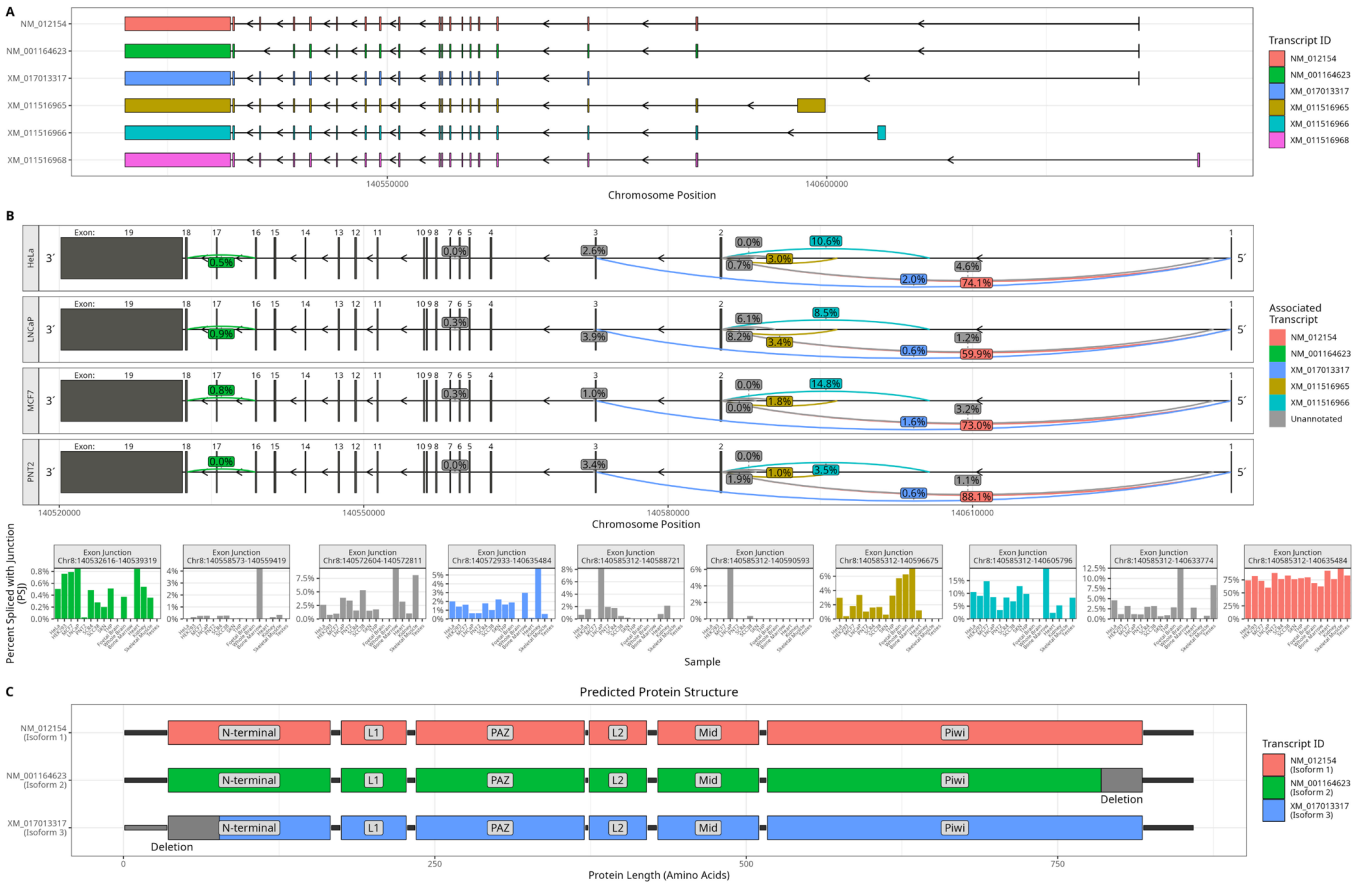

Figure 4.1. Alternative splicing events from RNA Capture-seq read data indicate multiple isoforms of Ago2 are expressed in humans. **A.** Refseq annotated transcripts for Ago2. Boxes show exons and lines show introns. **B.** Percent Spliced with Junction (PSJ) for each exon junction that varies across annotated transcripts, as well as the top 5 unannotated exon junctions by read count across all samples. PSJ was calculated as the proportion of reads including versus excluding an exon junction. Sashimi plots (top) for four representative cell lines show the position that each exon junction spans (curved lines) in relation to the wild-type transcript (NM_012154) and their respective PSJ values, coloured by the associated Ago2 transcript that each exon junction is unique to. Bar plots (bottom) show the PSJ values across exon junctions for all 16 cell lines and tissues. **C.** Shows the protein domains of NM_012154 (Isoform 1) and the location of amino acid deletions (dark grey) in the two splice variants selected for further study, NM_001164623 (Isoform 2) and XM_017013317 (Isoform 3).

4.4. Alternatively spliced Ago2 isoforms do not bind miRNAs as strongly as the canonical isoform

Ago2's primary function of gene regulation is mediated through its association with miRNAs, however prior studies have demonstrated that Argonaute can also bind other classes of small RNAs[293]. Compared to the canonical Ago2 isoform (Isoform 1), the two alternatively spliced isoforms we focused on have deletions in their protein coding sequences of the PIWI (Isoform 2) and N-terminal domains (Isoform 3). Both the PIWI and N-terminal domains have an important role in loading miRNAs during biogenesis, so we hypothesized that this would be altered in both isoforms and would change which small RNAs associate with them.

To study the small RNA binding properties of each isoform, DNA plasmids were transfected expressing each of the three Ago2 isoforms of interest. Each isoform's plasmid contained their respective protein coding sequence as well as a FLAG epitope at one of the terminal ends. The DNA plasmids were then transfected into HeLa cells, expressed for 24 hours, then immunoprecipitated with a FLAG antibody. RNA from the immunoprecipitate was subsequently isolated and sequenced for its small RNAs. The FLAG epitope was necessary to detect and isolate Argonaute isoforms as there are no known antibodies which can specifically target each of these isoforms. Western blots indicated that the FLAG-tagged Ago2 isoforms were strongly enriched after immunoprecipitation (Figure Appendix D).

Results from the sequencing data indicated that Ago2 Isoform 2 and 3 did not bind miRNAs as strongly as Isoform 1 (Figure 4.2). We first mapped all reads to known miRNAs from miRBase's miRNA database (v22.1), followed by sequential mapping of the remaining unmapped reads to ribosomal RNAs (rRNAs; Gencode's v35 and arb-silva's rRNA sequences), transfer RNAs (tRNAs; Gencode's v35 tRNA sequences), and finally other classes of small RNAs (scRNAs, snRNAs, snoRNAs, ribozymes, sRNAs, scaRNAs, vault RNAs; Gencode's v35). Due to the incomplete annotation of small RNA fragments derived from rRNAs (rRFs) and tRNAs (tRFs), we considered any mapping to a rRNA or tRNA as a positive match for that small RNA class. Most of the small RNAs from Isoform 1 were miRNAs (89.9-90.3%), which was consistent with miRNA-mediated gene regulation being its primary function (Figure 4.2). Although miRNAs were still the most abundant small RNA class in Isoform 2 (61.6-67.2%) and 3 (57.6-61.1%), miRNAs represented a much smaller proportion of small RNAs compared to Isoform 1. The order of abundance of small RNA classes was identical in all three Ago2 isoforms. Ribosomal RNAs were the second most abundant class, followed by snoRNAs, tRNAs, and snRNAs. The remaining small RNA classes collectively represented a sizeable pool in Isoform 2 (13.7-15.1%) and Isoform 3 (15.3-16.6%), compared to Isoform 1 (3.24-3.63%). There was a general increase in the relative proportion of all other small RNA classes in Isoform 2 and 3, implying that this reduction was likely due to a

reduced association with miRNAs as opposed to an increased association with another class of small RNAs (Figure 4.2B). These results were consistent with our hypothesis that deletions in the PIWI and N-terminal domains would lead to changes in miRNA loading efficiency and implied that cells which predominantly express these isoforms may have a general reduction in miRNA activity.

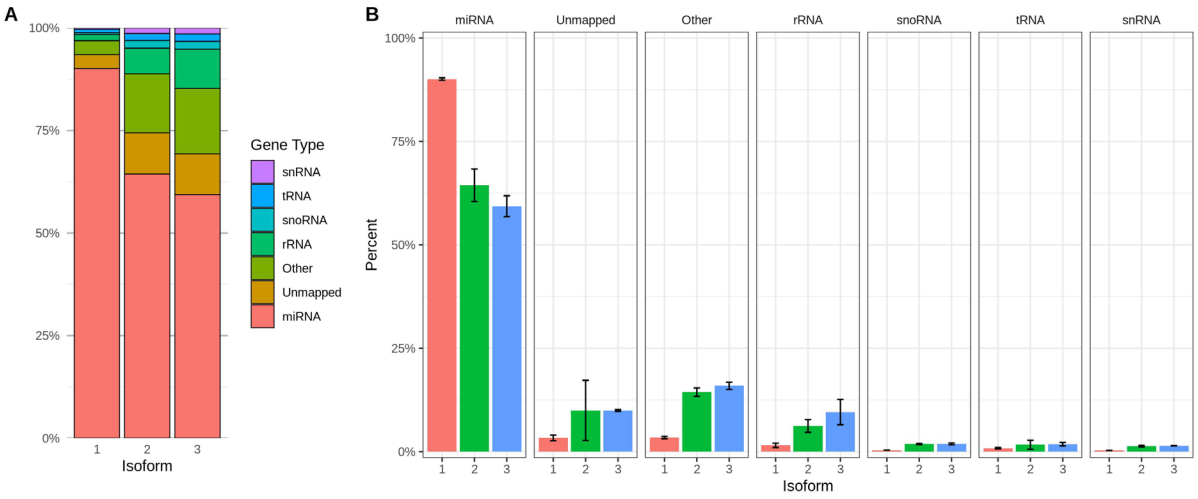

Figure 4.2. Proportion of small RNA classes immunoprecipitated from Ago2 isoforms in HeLa cells. Reads were sequentially mapped to miRNAs, rRNAs, tRNAs and all remaining small RNAs. The other small RNAs (scRNA, ribozyme, sRNA, scaRNA, vault RNA) were combined into the 'Other' category. **A.** Stacked bar chart showing relative abundance of each small RNA class for each isoform. **B.** Comparison of isoforms for each small RNA class. Error bars show standard error of the mean from two biological replicates. **miRNA:** MicroRNA, **rRNA:** Ribosomal RNA, **snoRNA:** Small Nucleolar RNA, **tRNA:** Transfer RNA, **snRNA:** Small Nuclear RNA, **scRNAs:** Small Cytoplasmic RNA, **sRNA:** Small RNA, **scaRNA:** Small Cajal body-specific RNA.

4.5. Ago2 isoforms have altered miRNA binding preferences

4.5.1. The relative abundance of certain miRNAs is altered in Ago2 isoforms

Our previous results revealed a reduction in miRNAs associated with Ago2 Isoforms 2 and 3. Following this, we investigated if this reduction in miRNA association applied to all miRNAs or if it was caused by a reduction in affinity towards specific miRNAs. We mapped and quantified miRNAs using miRbase's miRNA annotations (v22.1) and normalized the values to the total number of reads mapping to miRNAs (Figure 4.3). This way we could compare the relative abundance of each miRNA in different Ago2 isoforms, informing us of any preferential loading. We found a high correlation in miRNA abundance when comparing each Ago2 isoform, suggesting that the preferences towards miRNAs was generally similar across isoforms (Figure 4.3A-C). The strongest correlation was between Isoform 2 and 3 (r=0.95; Figure 4.3C), followed by Isoform 1 and 2 (r=0.92; Figure 4.3A), then Isoform 1 and 3 (r=0.85; Figure 4.3B). However, a small subset of miRNAs were differentially expressed when comparing isoforms (Table 4.2). This included 5 miRNAs that were differentially expressed between Isoform 1 and 2 (hsa-let-7b-3p, hsa-miR-25-5p, hsa-miR-27a-5p, hsa-miR-27b-5p and hsa-miR-365a-5p), and 7 miRNAs differentially expressed between Isoform 1 and 3 (hsa-let-7b-3p, hsa-miR-181a-3p, hsa-miR-181a-2-3p, hsa-miR-25-5p, hsa-miR-27a-5p, hsa-miR-27b-5p and hsa-miR-424-3p). All differentially expressed miRNAs had reduced abundance in Isoform 1, and four of these miRNAs (hsa-let-7b-3p, miR-25-5p, hsa-miR-27a-5p, hsa-miR-27b-5p) were common to both comparisons. Only one miRNA, hsa-miR-25-5p, was significantly different when comparing Isoform 2 and 3, which was identified in all comparisons of isoforms. The difference in miR-25-5p abundance was pronounced in each isoform, as it was only the 168th most abundant miRNA in Isoform 1, increasing to 54th in Isoform 2 and 9th in Isoform 3. For miR-27b-5p, this was only the 186th most abundant miRNA in Isoform 1 but increased to 79th in Isoform 2 and 22nd in Isoform 3. Together the results suggested that some Ago2 isoforms have preferences towards a small set of miRNAs, which may lead to a change in gene regulation for the targets of these miRNAs.

4.5.2. Alternative splicing of Ago2 changes miRNA strand selection preference

Published sequencing data on miRbase revealed that all the miRNAs identified from our differential expression analysis were typically the least expressed strand derived from their precursor, suggesting they were all passenger strands that are normally not favoured for selection during miRNA duplex loading. This led us to consider if strand selection for specific miRNAs was altered in these Ago2 isoforms, which could cause an increase in the measured relative abundance for certain miRNAs. To estimate strand selection preference, we quantified the number of reads mapping to the 5p and 3p strands for each miRNA precursor (Figure 4.4). We surmised that if the difference in abundance of miRNAs between isoforms was a consequence of a change in strand preference, the ratio of reads

mapping to the 5p and 3p strands would change, however if this observation was caused by precursor expression, we would expect there to be no change in their ratio. We included the top 50 expressed miRNAs in each Ago2 isoform, and excluded precursors only containing one strand, totalling 64 unique miRNAs (Figure 4.4). We found that most miRNA precursors strongly favoured a single strand, with a slightly greater number of miRNAs preferring 5p over 3p strands for selection. For a minority of miRNA precursors, including miR-92b, miR-423, and let-7d, the data suggested that both strands make a substantial contribution to the miRNA pool (Figure 4.4).

Strand preference was nearly identical for most miRNAs when comparing isoforms (Figure 4.4). However, a large difference was observed between isoforms for miR-25, miR-27b, miR-365a, as well as smaller changes in miR-221, miR-193b and let-7b. With miR-25, nearly all the reads mapping to this precursor in Isoform 1 were from the 3p strand (4.19% from 5p strand). However, in Isoform 2 the proportion of 5p miRNAs increased (34.4%) and in Isoform 3 the 5p strand became the dominant strand (78.8%). Likewise, for miR-27b, reads from this precursor were almost entirely from the 3p strand in Isoform 1, with only 1.15% of reads coming from the 5p strand. In Isoform 2 this increased to 12.2% from the 5p strand and in isoform 3 there was a larger change with a similar amount of reads from both strands (48.9% from 5p strand). For miR-365a, Isoforms 2 and 3 almost completely favoured the 5p strand (95.5% and 94.6% respectively) whereas isoform 1 had only 67.3% from the 5p strand. These observations may explain why some miRNAs were differentially expressed in the Ago2 isoforms we examined. Changes in strand selection probabilities would be expected to alter the abundance of each strand, and relatively minor shifts in strand preference for highly expressed miRNA precursors could significantly alter the abundance of their passenger strands. However, strand selection did not seem to be affected with all the differentially expressed miRNAs identified previously, such as miR-181a-3p and miR-424-3p, and another explanation is likely relevant for these miRNAs.

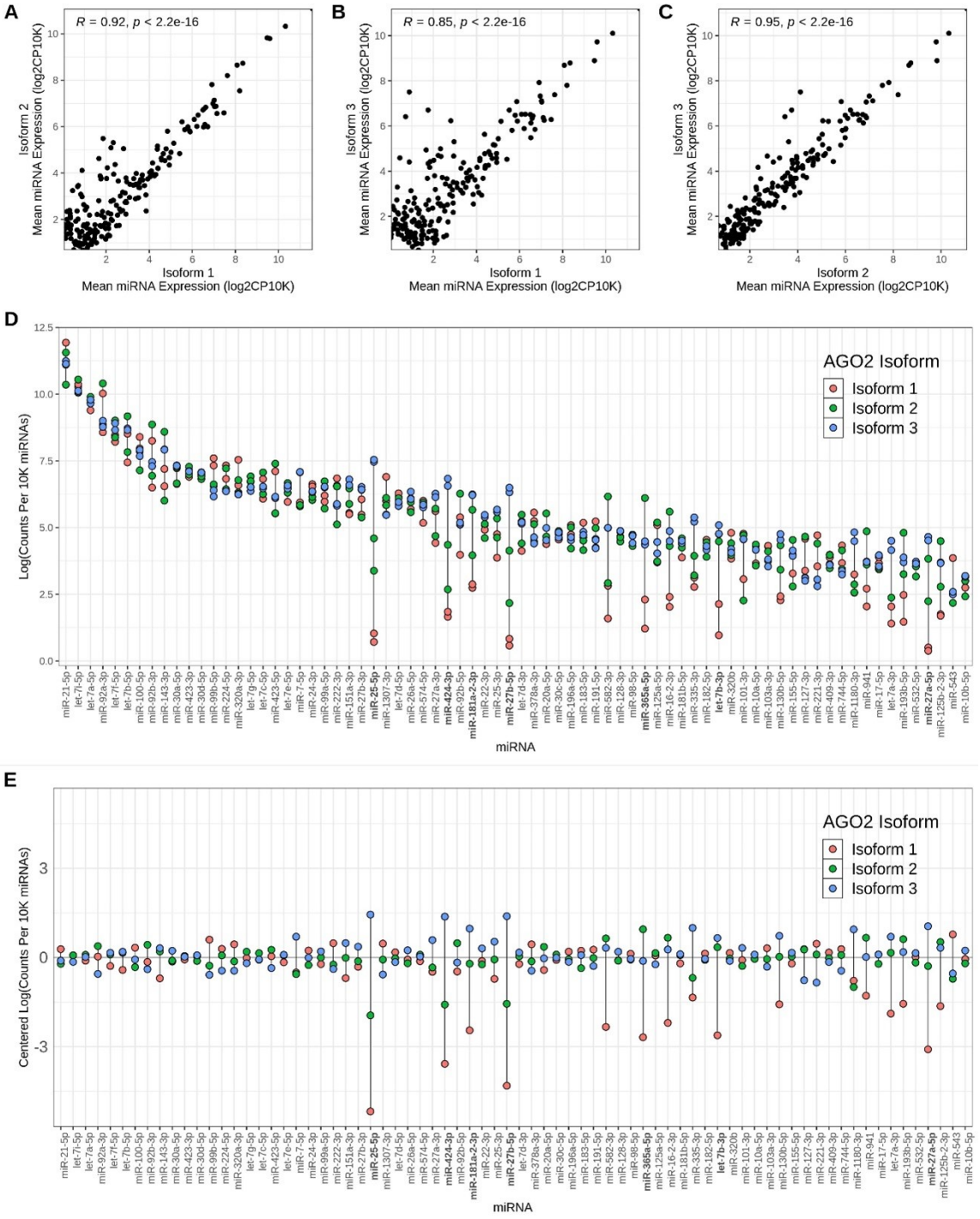

Figure 4.3. Comparison of miRNA expression between isoforms. **A.** Correlation of top 200 miRNAs according to mean expression (normalized to total miRNAs) across all samples, for Isoform 1 (x axis) and 2 (y axis). **B.** Isoform 1 (x axis) and 3 (y axis). **C.** Isoform 2 (x axis) and 3 (y axis). **D.** Shows miRNA expression in $\log_2$ (Counts per 10000 miRNAs) for all biological replicates of Ago2 isoforms. **E.** Shows mean Ago2 isoform expression in $\log_2$ (Counts per 10000 miRNAs), centered for each miRNA. miRNAs that are differentially expressed in at least one isoform comparison are shown in bold.

miRNA	log2FoldChange	P-value	P-value (adj.)	Isoform Comparison
hsa-let-7b-3p	-3.11603	4.27E-07	1.43E-05	Isoform 1 vs Isoform 2
hsa-let-7b-3p	-3.3647	4.76E-08	1.79E-06	Isoform 1 vs Isoform 3
hsa-miR-181a-2-3p	-3.15678	1.05E-06	2.84E-05	Isoform 1 vs Isoform 3
hsa-miR-181a-3p	-4.07333	6.79E-08	2.44E-06	Isoform 1 vs Isoform 3
hsa-miR-25-5p	-3.94528	7.55E-11	1.57E-08	Isoform 1 vs Isoform 2
hsa-miR-25-5p	-7.29656	2.05E-33	1.77E-30	Isoform 1 vs Isoform 3
hsa-miR-25-5p	-3.35127	3.01E-08	5.25E-05	Isoform 2 vs Isoform 3
hsa-miR-27a-5p	-4.18793	4.58E-10	6.34E-08	Isoform 1 vs Isoform 2
hsa-miR-27a-5p	-5.49585	2.74E-16	5.93E-14	Isoform 1 vs Isoform 3
hsa-miR-27b-5p	-3.62575	4.29E-07	1.43E-05	Isoform 1 vs Isoform 2
hsa-miR-27b-5p	-6.54556	6.68E-20	2.89E-17	Isoform 1 vs Isoform 3
hsa-miR-365a-5p	-3.64105	2.26E-08	1.34E-06	Isoform 1 vs Isoform 2
hsa-miR-424-3p	-4.97974	3.24E-14	5.59E-12	Isoform 1 vs Isoform 3

Table 4.2. Differentially expressed miRNAs when comparing three isoforms of Ago2. Includes miRNAs with an adjusted P-value less than 0.05.

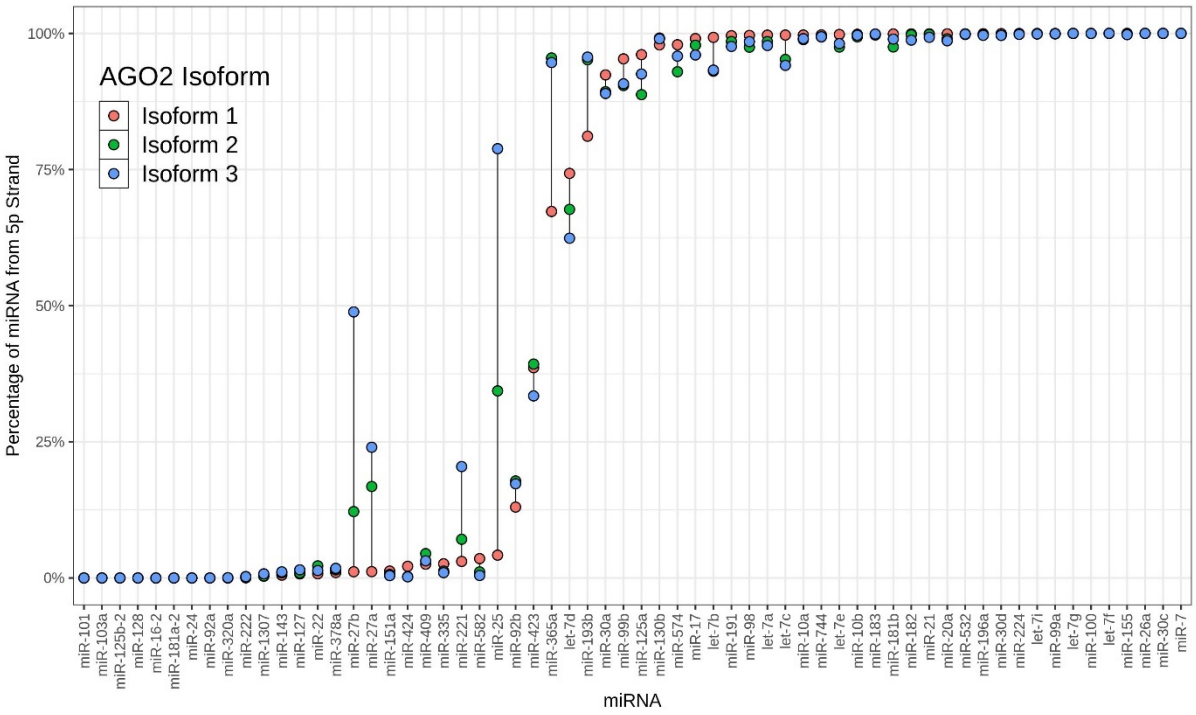

Figure 4.4. Isoforms show strand selection bias in specific miRNAs. For miRNA precursors which express two strands (5p and 3p), the percentage of reads from the 5p strand were calculated to estimate strand selection preference in each Ago2 isoform.

4.6. The 5′ nucleotide preference during strand selection is altered in Ago2 isoforms

We then investigated features of the miRNA precursor duplexes to identify what may be influencing the change in strand preference for different Ago2 isoforms. Strand selection is known to be determined by two primary factors – a 5′ nucleotide preference in the miRNA duplex and the relative thermodynamic stability of the terminal nucleotides on each side of the miRNA duplex. Previous studies have shown that the first base at the 5′ end of each strand influences the likelihood of it being selected by Argonaute, with bases prioritized in the following order – Uridine > Adenine > Cytosine > Guanine[304]. Additionally, the miRNA strand whose 5′ end lies on the least thermodynamically stable side of duplex, determined by the last 5 nucleotides, is often favoured for selection. We considered if the 5′ nucleotide preference was altered in the isoforms of Ago2, which may explain why some miRNA strands were more favoured.

Base selection preference was determined by estimating the probability of each base being selected over other bases using our data. For this we considered that each miRNA-mapped read represents one base selection event, where it's first nucleotide (5′ end) was selected over the first nucleotide of the opposite strand. The opposite strand's sequence was predicted by examining its canonical miRNA sequence in miRbase, as it is not possible to know the actual sequence of the opposite strand which is discarded after loading. As 5′ isomiRs were present at a low frequency in our isoform data (8.6-10.4% of total miRNAs), we expected this prediction to be accurate for most events. We also excluded all reads annotated as 5′ isomiRs as we do not know if the alteration at the 5′ end was introduced before or after miRNA duplex loading and strand selection.

We found that each isoform of Ago2 generally adhered to the same base priorities and were similar to what has been established in previous studies (Figure 4.5). The only exception was in duplexes with Guanine and Adenine, where Guanine was slightly more favoured. Uridine was the most strongly favoured base over all other bases, where it was more likely to be selected over Adenine (58.7-68.3%), Cytosine (96.1-99.0%), and Guanine (98.7-99.2%). Adenine was strongly favoured over Cytosine (80.1-87.5%) although was slightly less favoured over Guanine (39.8-49.1%). In duplexes with Cytosine and Guanine on each end, Cytosine was strongly favoured (76.4-84.9%).

Next, we separated each type of base selection event by their miRNA precursor to investigate if this base priority held true for each precursor's duplex (Figure 4.6). We found a large variation in base preference between miRNA duplexes, and although they appeared more likely to follow these rules there were clear examples which did not, such as miR-31, where Adenine was selected over Uridine 99.2%-99.5% of the time (Figure 4.6B). For each base selection event, we divided miRNA duplexes by which bases was more probable and found that Uridine was the preferred base in 81.0% (68/84) of

duplexes for Isoform 1. This number decreased to 72.6% (61/84) in Isoform 2 and 70.2% (59/84) in Isoform 3, suggesting the preference for Uridine remained but was not as strong in these isoforms (Figure 4.6). Adenine was preferred base in 55.6% (30/54) of duplexes in Isoform 1, to 61.1% (33/54) in Isoform 2 and 64.8% (35/54) in Isoform 3. Cytosine was preferred in 39.6% (21/53) for Isoform 1, 47.2% (25/53) for Isoform 2 and 45.3% (24/53) for Isoform 3. Finally, Guanine was similar between isoforms and was preferred in 68.0% (17/25) of duplexes for each isoform. The increased number of miRNA duplexes preferring Adenine and Cytosine when comparing Isoform 2 and 3 to Isoform 1 was primarily due to the decrease in preference for Uridine (Figure 4.6). We noted for most miRNA duplexes where one base was very strongly favoured (>99%) in one Ago2 isoform, base preferences did not change. The difference in base preference frequency between isoforms was more apparent in duplexes where base selection was already ambiguous, including miR-92a-1 (Figure 4.6A; U>A), miR-132 (Figure 4.6A; U>A), miR-30b (Figure 4.6A; U>C), miR-485 (Figure 4.6B; A>G), and miR-365b (Figure 4.6B; A>U).

Our data is consistent with previous assumptions that base rules influence strand selection, however each duplex likely harbors unique features (for example relative thermostability) which also contribute to strand selection and may override the 5´ base preference. Although Ago2 Isoforms 2 and 3 seem to follow the same principles of base preference, the strength of these preferences is reduced, and this may have a large effect on which strands are loaded into Argonaute for a subset of miRNA duplexes. Evidently, this can change the pool of active miRNAs within a cell and would have consequences for the regulation of miRNA targets. Together these results implicate alternative splicing of Argonaute as a novel mechanism of gene regulation through the modulation of miRNA strand selection.

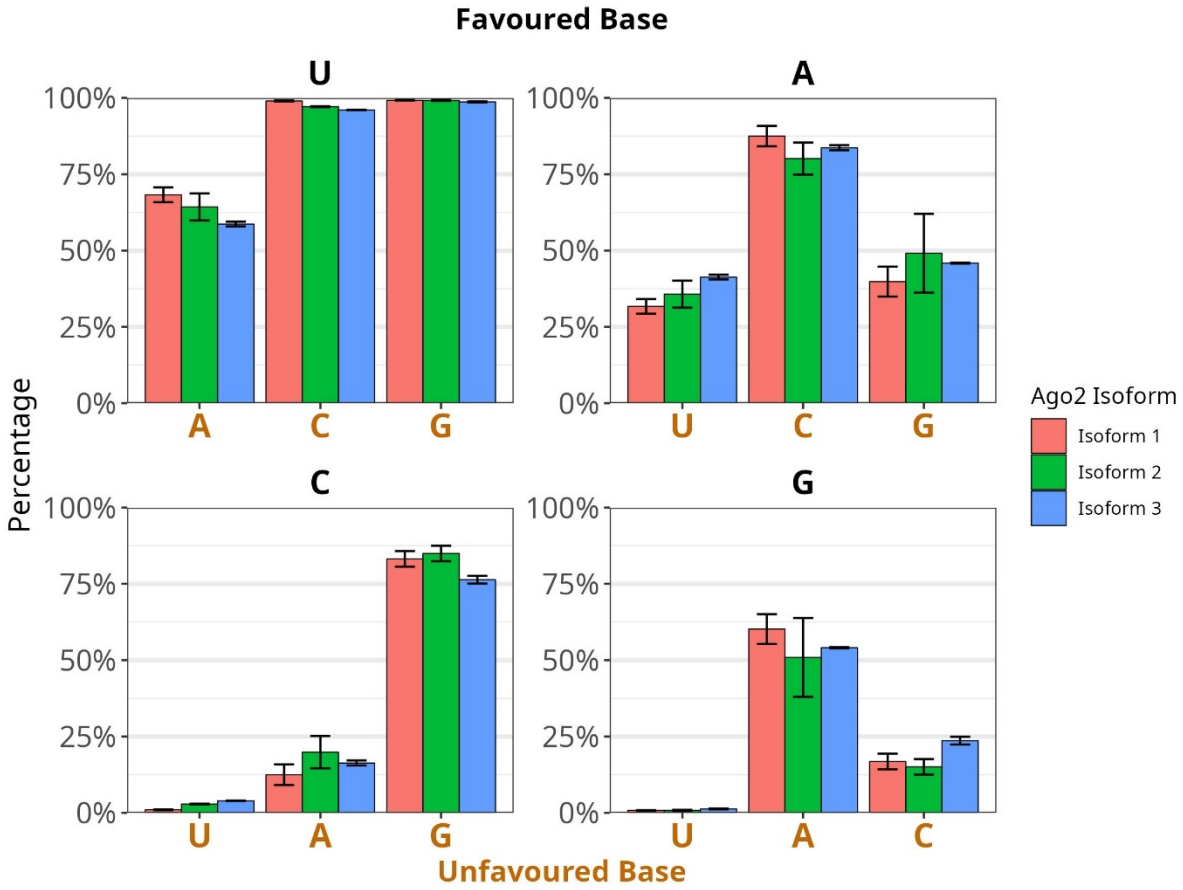

Figure 4.5. Base preferences from all miRNA duplexes for each pair of bases. Shows the percentage of cases where one base is favoured (top, black) over another (bottom, brown) for each combination of bases. Favoured bases are determined by considering the first nucleotide of each read (guide strand) and comparing it to the first nucleotide of the predicted passenger strand. Error bars show standard error of the mean from two biological replicates.

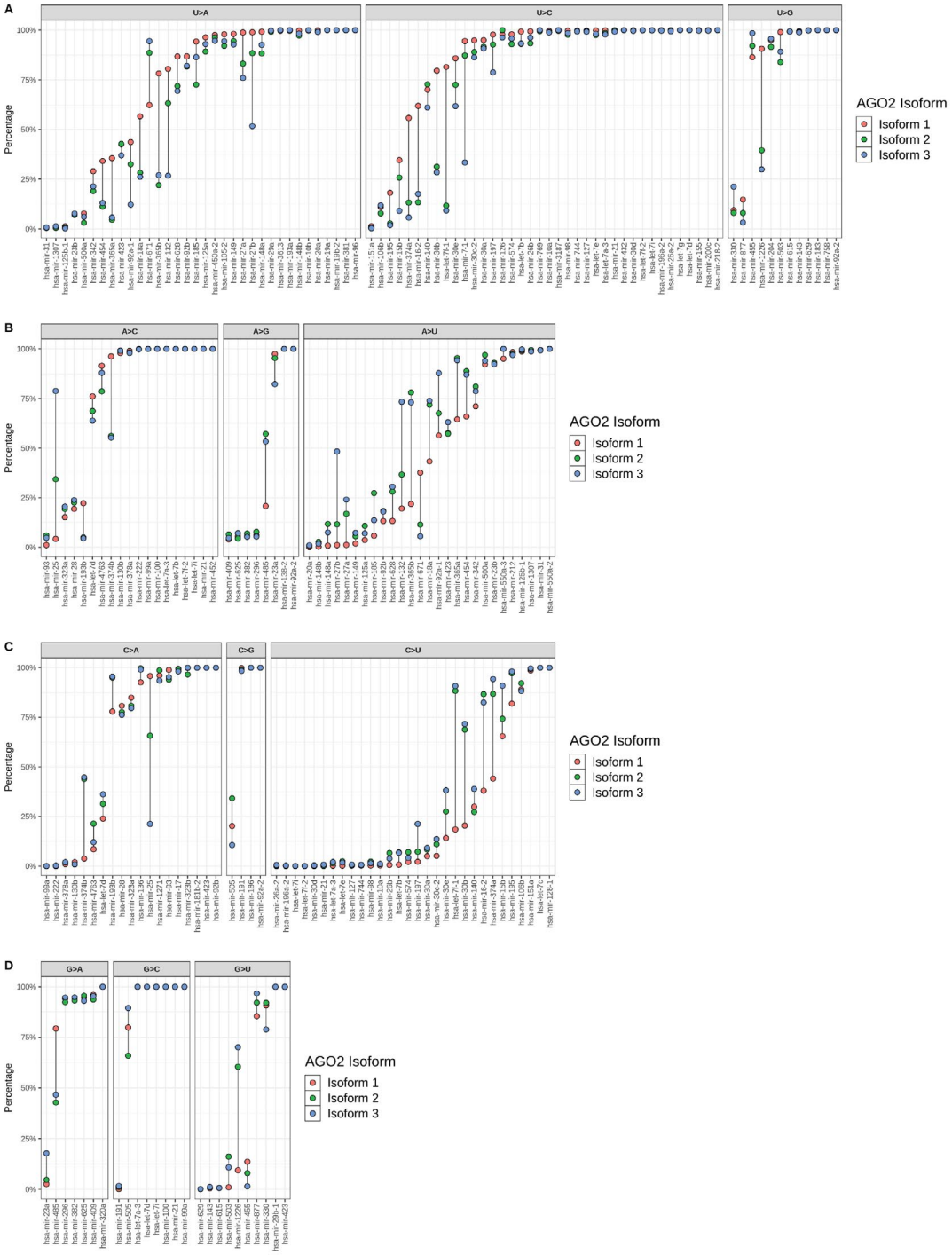

Figure 4.6. Base preferences of miRNA duplexes from each precursor miRNA. Shows first 5´ nucleotide of guide strand being favoured over their predicted passenger strand's 5´ nucleotide (guide>passenger) for **A.** Uridine, **B.** Adenine, **C.** Cytosine, **D.** Guanine.

4.7. Ago2 isoforms have altered isomiR associations

Our results revealed a change in 5′ nucleotide preference for Ago2 isoforms that affected strand selection. We hypothesized that the length or nucleotide alterations of isomiRs may also change how they interact with the Ago2 isoforms and could lead to differences in their relative abundance in RIP-seq data.

We first considered if Ago2 isoforms were biased towards miRNAs of certain lengths. Although we observed an increase in the proportion of 23 nucleotide long miRNAs for Ago2 Isoform 1, there was a high variance between replicates and no significant difference was shown between the isoforms for any length (Figure 4.7).

Following this, we investigated if the Ago2 isoforms had a bias towards certain isomiR types by examining the proportion of miRNAs annotated as each isomiR type (Figure 4.8). All three Ago2 isoforms had a higher proportion of isomiRs compared to canonical miRNAs. However, there were minor differences between Ago2 isoforms, where Isoform 1 had the lowest proportion of isomiRs at 61.3%, Isoform 2 at 64.3%, and the highest proportion of isomiRs in Isoform 3 (65.4%). Isoform 1 had the highest proportion of 5′ variants (10.5%), compared to Isoform 2 (8.7%) and Isoform 3 (8.6%). This difference was reflected in both longer and shorter 5′ variants. Substitutions were highly similar in all isoforms, at 8.8% for Isoform 1, 8.8% for Isoform 2 and 8.7% for Isoform 3. For 3′ templated variants, the highest proportion was observed in Isoform 3 (50.5%), followed by Isoform 2 (48.8%), and finally Isoform 1 (47.0%). However, this was only true for the shorter 3′ templated variants, whereas the longer 3′ templated variants had a higher proportion in Isoform 1 (17.9%) compared to Isoform 2 (17.2%) and Isoform 3 (15.9%). We also considered 3′ non-templated variants for all four bases and found that Isoform 2 and Isoform 3 had a higher proportion of miRNAs with non-templated Adenine, Uridine and Cytosine additions compared to Isoform 1. Although much rarer, the opposite was true for 3′ non-templated Guanines which were higher in Ago2 Isoform 1 compared to Isoform 2 and Isoform 3.

To investigate this further we compared isomiR types for each miRNA between Ago2 isoforms. The change in proportion of isomiRs and isomiR types for individual miRNAs (Figure 4.9-4.12) frequently reflected what was observed with the total miRNAs (Figure 4.8). For example, a higher proportion of total isomiRs, 3′ templated, 3′ non-templated Adenine, 3′ non-templated Uridine, and 3′ non-templated Cytosine variants were observed for Isoform 1 compared to both Isoform 2 and 3 in most miRNAs. Interestingly a small set of miRNAs had a reverse trend to the other miRNAs. Many of these miRNAs were identified in the previous section, which may have different interactions with the Ago2 isoforms during strand selection. For example, most miRNAs had a modest decrease in the proportion

of isomiRs (Figure 4.9A) when comparing Isoform 1 with Isoform 2 and 3. However, several miRNAs including miR-424-3p, miR-27b-5p, miR-365a-5p, miR-193b-5p, miR-16-2-3p and miR-25-5p had a higher proportion of isomiRs in Isoform 1 compared to Isoform 2 and Isoform 3. Additionally, the proportion of 3´ templated variants was higher in Isoform 1 for miR-424-3p, miR-27b-5p, let-7a-3p, miR-365a-5p, miR-193b-5p and miR-25-5p compared to Isoform 2 and 3, but lower in Isoform 1 for all other miRNAs (Figure 4.11A).

The data suggests that Ago2 isoforms have low bias towards certain types of isomiRs. Although in most cases the impact on the overall pool of loaded miRNAs is modest, there is further evidence of distinct interactions between Ago2 isoforms and some miRNAs and the effect on gene regulation for these miRNA targets may be significant.

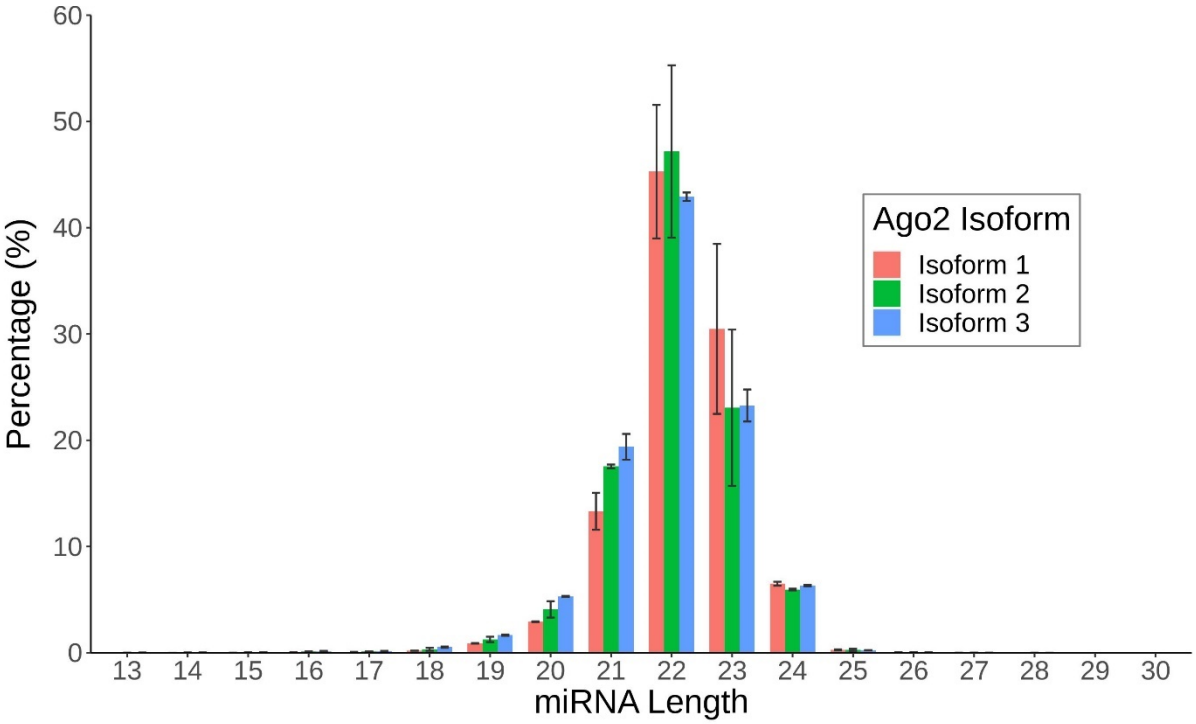

Figure 4.7. Comparison of the size of miRNAs associated with the three Ago2 isoforms. Error bars show standard error of the mean from two biological replicates.

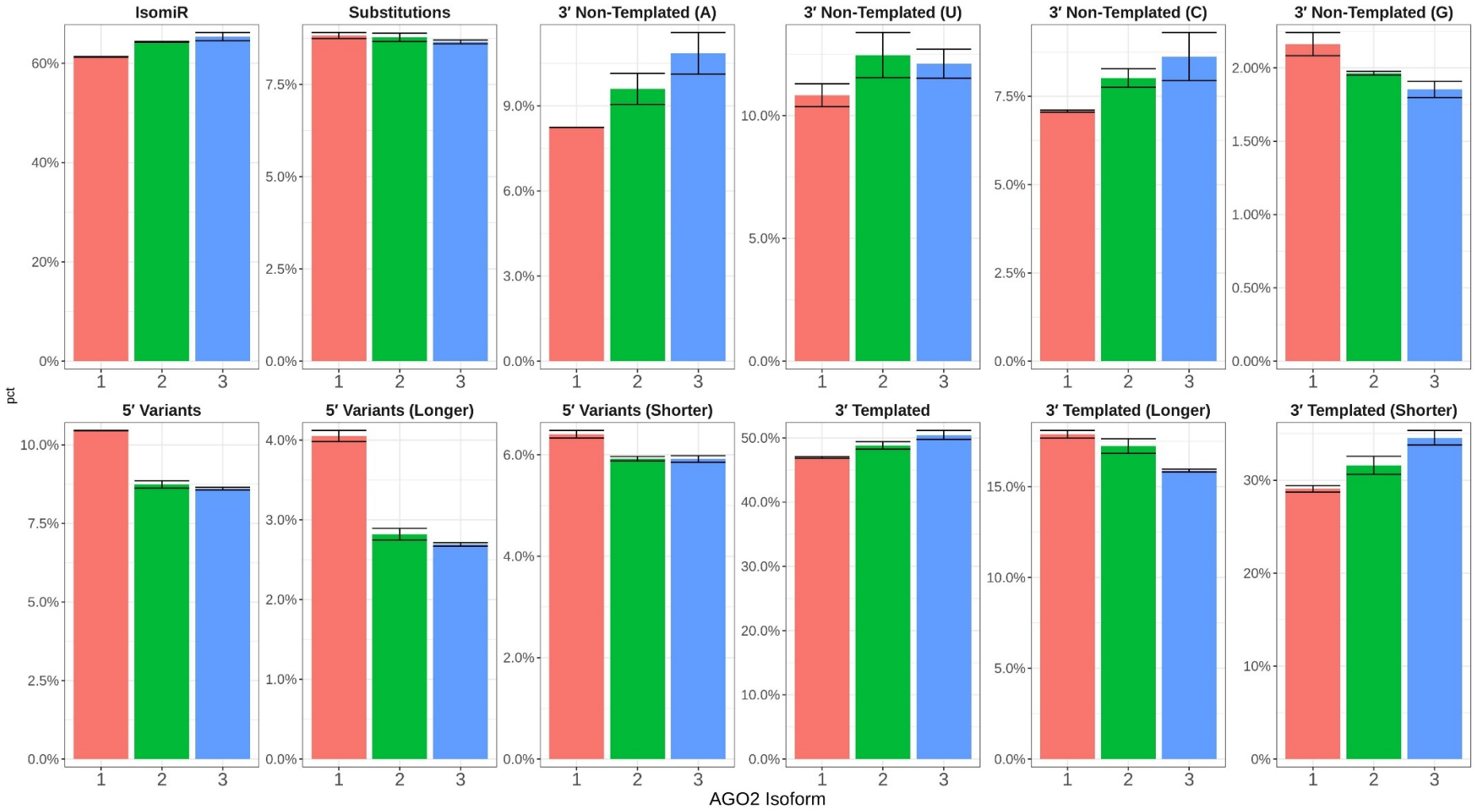

Figure 4.8. IsomiR types associated with Ago2 isoforms. Includes Isoform 1 (red), Isoform 2 (green), and Isoform 3 (blue). 5′ variants, 3′ templated and 3′ non-templated isomiR types were also subdivided into those longer or shorter than the canonical miRNA. Error bars show standard error of the mean from two biological replicates.

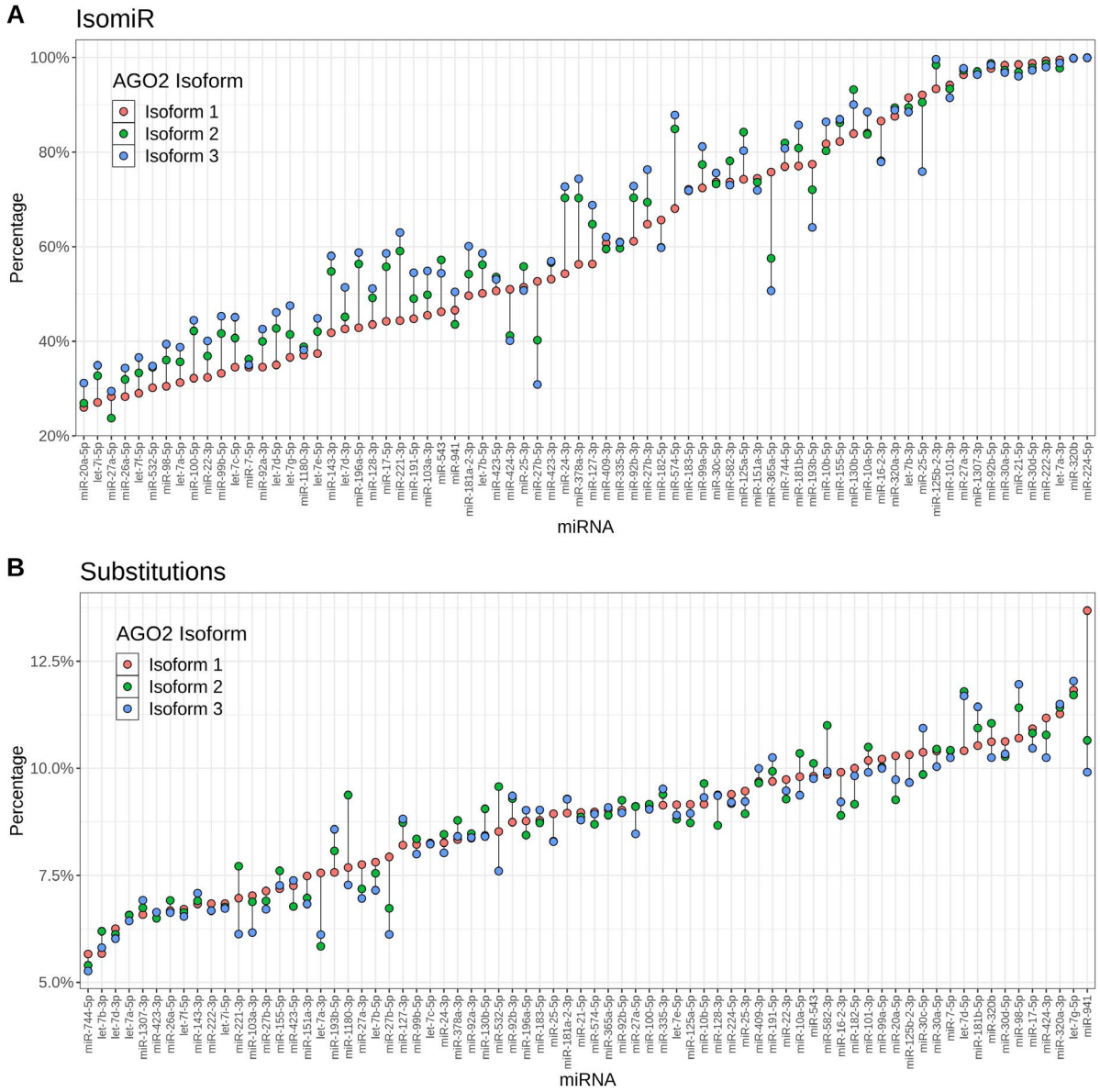

Figure 4.9. Proportion of miRNAs associated with Ago2 isoforms, annotated as **A.** isomiRs (non-canonical miRNAs) and **B.** substitution variants. Isoform 1 is shown in red, Isoform 2 in green and Isoform 3 in blue.

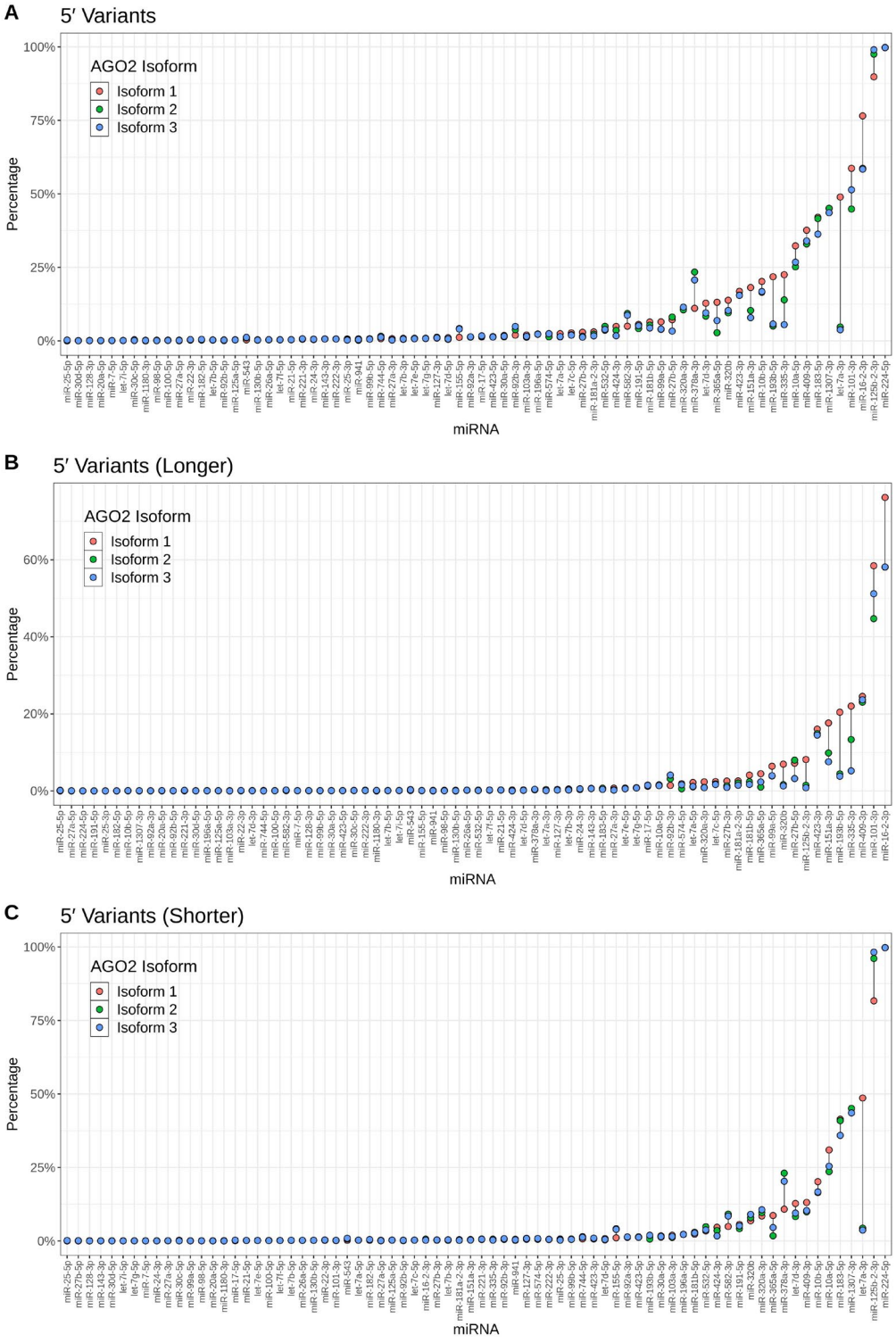

Figure 4.10. miRNAs associated with Ago2 isoforms annotated as 5′ variants. **A.** Shows percentage of each miRNA that are 5′ variants, **B.** 5′ variants longer than the canonical sequence, and **C.** 5′ variants shorter than the canonical sequence. Isoform 1 is shown in red, Isoform 2 in green and Isoform 3 in blue.

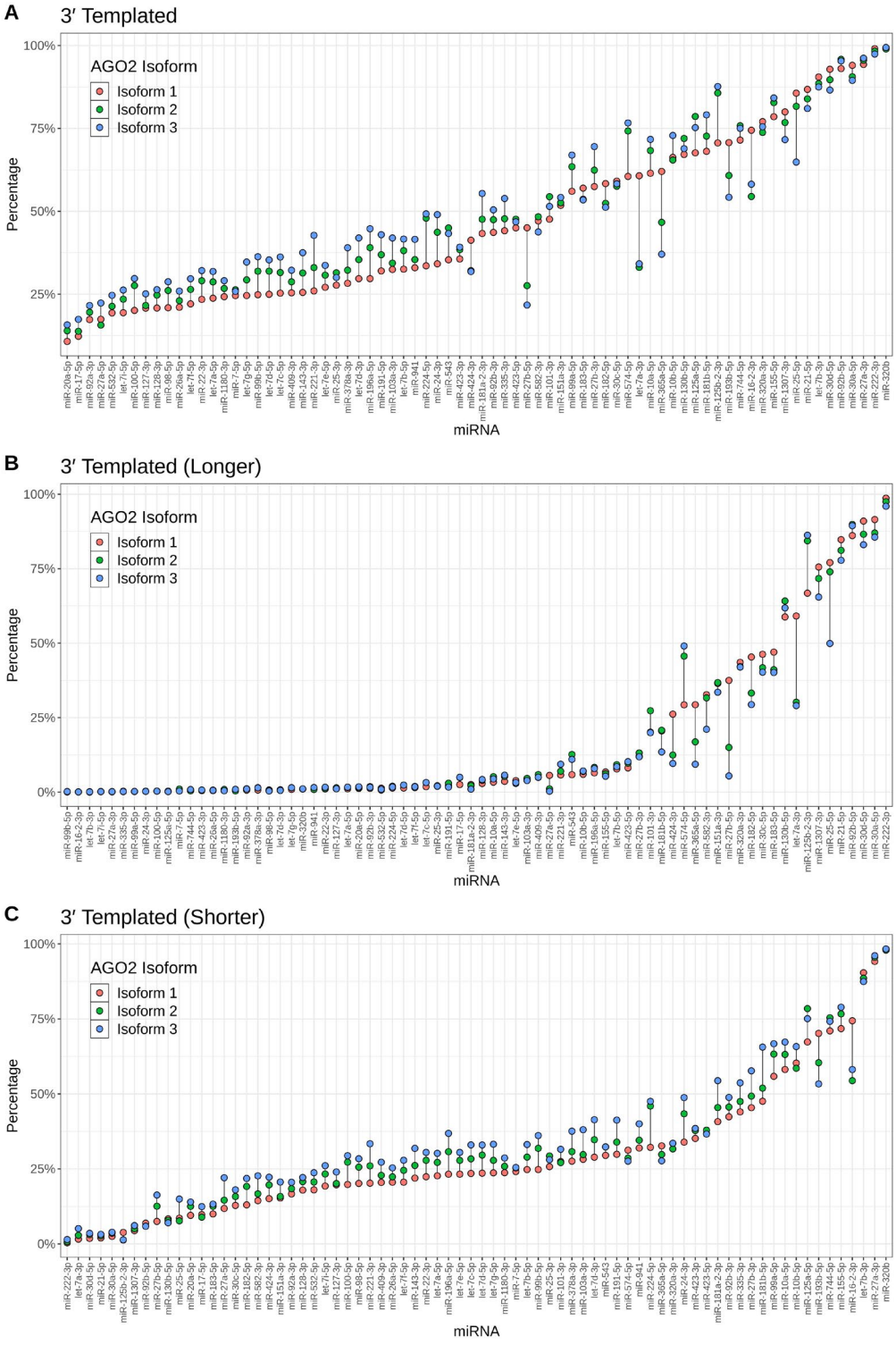

Figure 4.11. miRNAs associated with Ago2 isoforms annotated as 3′ templated variants. **A.** Shows percentage of each miRNA that are 3′ templated variants, **B.** 3′ variants longer than the canonical sequence, and **C.** 3′ variants shorter than the canonical sequence. Isoform 1 is shown in red, Isoform 2 in green and Isoform 3 in blue.

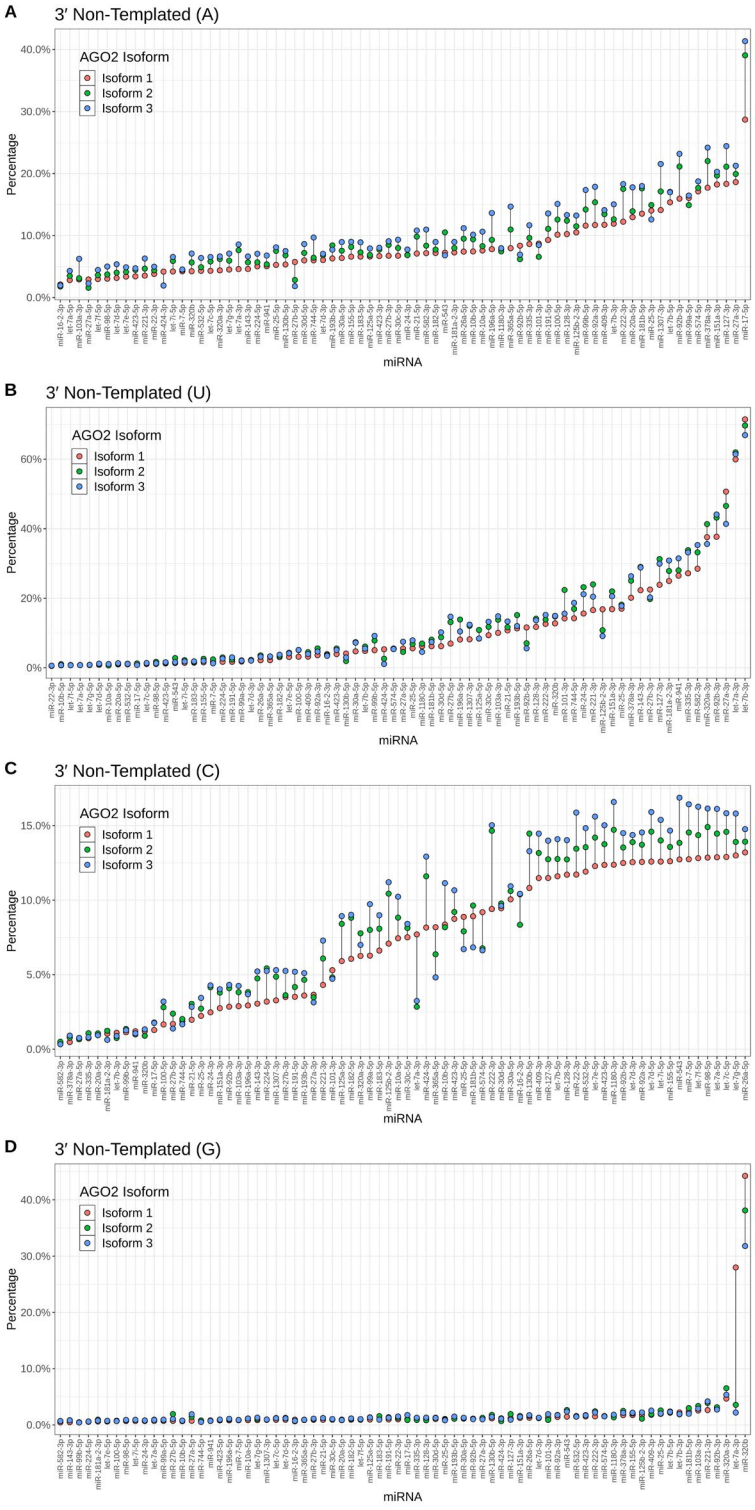

Figure 4.12. miRNAs associated with Ago2 isoforms annotated as 3′ non-templated variants. Shows percentage of each miRNA that contains a **A.** 3′ non-templated Adenine, **B.** 3′ non-templated Uridine, **C.** 3′ non-templated Cytosine, **D.** 3′ non-templated Guanine. Isoform 1 is shown in red, Isoform 2 in green and Isoform 3 in blue.

4.8. Discussion

Argonaute proteins encoded by different genes can discriminate between some species of miRNAs[293,309]. However, the biological relevance of alternatively spliced variants of Argonaute are poorly understood. Here, we identified three isoforms of Ago2 which are expressed in human cells and show that these isoforms interact with miRNAs differently, suggesting distinct roles within the miRNA pathway. We found a reduction in miRNA abundance in Ago2 Isoforms 2 and 3, described above, which suggest that for cells with extensive splicing and/or expression of these isoforms the miRNA pathway may exert a weaker regulatory effect on their gene targets due to a reduction in the number of RISC-associated miRNAs. Although this assumption is intuitive, the deletions in the PIWI and N-terminal domains of Isoforms 2 and 3, respectively, may not just affect miRNA selection and loading but also the regulatory activity of the RISC. Preliminary data from our lab has confirmed a reduction in let-7a and miR-21 association for Ago2 Isoform 2 and 3 compared to Isoform 1 using RT-qPCR, as well as a reduced capacity to form a RISC (Bajan, unpublished data). Collectively, this evidence supports a reduction in miRNA-mediated gene regulation for Ago2 Isoforms 2 and 3.

However, this reduced role may not apply to all miRNAs. Several miRNAs, including miR-25, miR-27a, miR-27b and miR-365a were more abundant in the RIP-seq data for Ago2 Isoform 2 and 3. We attributed this to a change in strand preference between the 5p and 3p strands of the miRNA duplex during loading. Although both the PIWI and N-terminal domains participate in miRNA loading, it was surprising to see similar changes in miRNA selection for Ago2 Isoform 2 and 3 compared to Isoform 1. One explanation for this is that both deletions may interfere with the MID domain's rigid loop structure that affords Argonaute its strand selection bias[304]. Unwinding of the duplex is initiated by the N-terminal domain, which pries the duplex open and positions the 5´ terminal base of the potential guide strand at the interface of the MID and PIWI domain[15]. Deletions in the PIWI and N-terminal domain may alter how the duplex is positioned within the pre-RISC, reducing the bias this loop structure has on base preference. An alternative explanation is that both deletions may alter Ago2's interaction with protein cofactors which assist it during RISC formation and contribute to strand selection.

4.8.1. Limitations of sequencing data for the investigation of Argonaute strand selection

RIP-seq uses antibodies to capture target proteins and the RNAs associated with them[310]. It is an assumption that when targeting Argonaute, the miRNAs sequenced through this method would be guide strands that were loaded into an Argonaute protein, as passenger strands are discarded and degraded soon after RISC formation[67]. Because of this, our methodology for determining strand selection preference required a prediction of the passenger strand for each captured guide miRNA.

We focused on the first 5´ nucleotide of each miRNA strand as this is known to have a significant influence on strand selection. Previous studies have shown that Argonaute generally prefers bases in the order of Uridine > Adenine > Cytosine > Guanine[299,304]. However, one study that examined 832 miRNA precursors annotated in miRbase with corresponding sequencing data, found only 429 (51.6%) followed these base preference rules[299]. Approximately 75% of the miRNAs annotated in miRbase could be explained by either the 5´ nucleotide order or their relative thermostability[299]. Our data was generally consistent with this base order, except that Guanines were more likely to be selected over Adenines. We found that a high proportion of the reads which indicated Guanine was selected over Adenine were from a single miRNA, miR-409-3p, and may have skewed this result. Clearly, a multitude of features in the miRNA duplex are likely to play a role in strand selection.

Finally, strand selection is not the only factor which determines miRNA abundance, and we cannot rule out any changes in miRNA stability or degradation that would alter the ratio of each strand's abundance.

4.8.2. Ago2 isoforms may interact with isomiRs to fine-tune miRNA strand selection

A substantial proportion of miRNAs are expressed as isomiRs, and many are produced in the early stages of miRNA biogenesis prior to RISC formation. In principle, 5´ variants can change the terminal bases of each strand, and all types of isomiRs may change the thermodynamic properties of a duplex[311]. Indeed, studies have shown how isomiRs produced from alternate Dicer processing and 3´ uridylation can affect strand selection[312,313]. Our evidence suggests that the isoforms of Ago2 that we studied differentially interact with isomiRs, so it is possible that the alternative splicing of Argonaute and the production of isomiRs are two mechanisms that work in concert to fine-tune the frequencies for each miRNA strand's selection. Although not included in this study, RNA modifications such as Adenine to Inosine (A-to-I editing) and chemical modifications (e.g locked nucleic acids), can also lead to changes in miRNA strand selection and it remains to be seen if Argonaute isoforms are distinguished in their interaction with these modifications[314,315].

4.8.3. Implications of Argonaute isoforms altering miRNA activity in normal and cancerous cells

Most studies assume that cytoplasmic miRNA levels are representative of what is loaded into RISCs, even though previous studies have shown that this is not always true[293]. Although we found many similarities in the types of miRNAs each Ago2 isoform associated with, there was notable examples of miRNAs which were nearly absent in one Ago2 isoform but strongly associated with another. This was most evident with miR-25-5p, which was the 9th most abundant miRNA in Ago2 Isoform 3 yet only the 168th most abundant in Isoform 1. Therefore, it is expected that the genes targeted by the miRNA pathway can change depending on the expression of Ago2 isoforms. We included several cancer cell

lines in our study, including MCF7 (breast cancer), LNCaP (prostate adenocarcinoma), SCC38 (squamous cell carcinoma), and THP-1 (monocytic leukemia), and found evidence of differential Ago2 splicing and isoform expression. Dysregulation of Ago2 splicing may contribute to disease states by altering the activity or function of oncogenic or tumour suppressive miRNAs. Many of the miRNAs we found associated with specific Ago2 isoforms, including let-7b-3p, miR-25-5p, miR-27b-5p have documented roles in regulating cancer pathways[316–321]. However further research is needed to identify if Ago2 splicing is perturbed in any cancers and if it is directly involved in altering the activity of these miRNAs.

In conclusion, our study identified alternative splicing of Argonaute as a novel mechanism which regulates how miRNAs are loaded into the RISC and may alter miRNA activity in normal and cancerous cells.

5. Final discussion

5.1. A multitude of mechanisms regulate the miRNA pathway

Gene regulation is a fundamental biological process that ensures cells can function properly throughout their entire life cycle[322]. Multicellular organisms, including humans, have evolved many different mechanisms for gene regulation, such as epigenetics, transcription initiation and alternative splicing, the miRNA pathway, and post-translational protein modifications, all to control gene expression spatially and temporally with a high degree of precision[196,322–324]. As research continues to show, each of these gene regulatory mechanisms is immensely complicated and can be modulated to serve the specific requirements of each cell[196,322–324]. Emerging evidence suggests that nearly every aspect of the miRNA pathway is tightly regulated[196,325]. At each stage of biogenesis, the immature miRNAs interact with proteins such Drosha, Dicer, and many auxiliary proteins, to regulate the pool of mature miRNAs and ultimately influence their function[1,51,53].

We identify a novel intersection between two gene regulatory mechanisms, the miRNA pathway and alternative splicing, which influence the selection of miRNAs by controlling the expression of various Argonaute isoforms. As mentioned previously, there is already evidence that alternative splicing regulates miRNA biogenesis through isoforms of Drosha and Dicer, however it is unclear how widespread their expression is across human tissues[24,26]. We identified at least 2 isoforms, which can both be expressed as the dominant isoform in various human tissues and cells. Our evidence suggests miRNA duplex loading is affected in some isoforms of Ago2, which can alter the pool of miRNAs that are functionally active. Interestingly, biases in miRNA selection appear to be restricted to a subset of miRNAs. A likely explanation for this, is that if Argonaute isoforms bound completely different miRNAs, then cells may not be able to tolerate the drastic changes in miRNA activity that would be induced by modulating their expression. We suspect that other factors associated with the miRNA duplex, such as the relative thermodynamic stability of its ends, allows the activity of some miRNAs to remain stable when Ago2 isoform expression changes, while others are more greatly impacted.

Another key finding in our study was evidence that isomiR processing is regulated cell autonomously. It is possible that a combination of changes in protein expression, as well as alternative splicing, contribute to the unique isomiR profiles that different cell types maintain. Dicer, Drosha, and many other proteins involved in the miRNA pathway may have splice variants that favour the processing of certain isomiRs[24,26]. As our work investigating isomiR expression using single cell sequencing is the first of its kind, it is likely that future research will continue to unearth a complex regulatory layer that has not been appreciated to date.

5.2. Different mechanisms of miRNA regulation may be perturbed in cancers

miRNA dysregulation is a common feature in nearly all cancers, therefore a better understanding of all the mechanisms regulating the miRNA pathway have potential implications in cancer research.

To investigate if miRNAs contribute to heterogeneous gene regulation, we analysed miRNA and isomiR expression from glioblastoma single cells. Our study identified two miRNA clusters, the Dlk1-Dio3 and miR-224/452 clusters, which are differentially expressed in two cell subpopulations from a glioblastoma primary cell culture[326]. As paired single cell small RNA and RNA sequencing data was unavailable in glioblastoma, we integrated data from other forms of sequencing, including bulk RNA-seq and small RNA-seq, as well as single cell RNA-seq, to predict the activity of these miRNAs in different cell states. A key aspect of this approach involved analysing the expression of miRNA targets, however in this study we highlighted several challenges including poor performance of target prediction algorithms and the low signal-to-noise ratio in single cell data[247,254]. Despite this, we found evidence of an association between the identified miRNA clusters and known glioblastoma cell states, in particular the NPC and MES states, supporting a potential role for the miRNA pathway in cell state regulation. We did not find evidence of altered isomiR processing between the glioblastoma cells in this study, however our sample size was small and only included three primary cultures and one cell line. Additionally, more extensive miRNA and isomiR heterogeneity may be observed in cells that are sourced directly from fresh tumours, as they exist within a more complex microenvironment[157].

We then expanded our analysis to include single cell miRNA sequencing data which included cell types from other cancers such as leukemia and hepatocellular carcinoma. Our work found evidence that individual cells precisely regulate the expression of their isomiRs and suggests some of these isomiRs are likely to possess distinct regulatory functions with respect to their miRNA targets. We identified several miRNAs, expressed as multiple isomiRs, which are likely to have altered regulation of their targets in both leukemia and hepatocellular carcinoma cells. Some of these miRNAs are highly expressed in cells derived from their normal tissues and it is possible their dysregulation may be involved in either disease[327–330]. For example, miR-92a-3p is one of the most highly expressed miRNAs in leukemia and other blood-derived cells, and within our study we found nearly all this miRNA was expressed as 3´ non-templated variants that were less active in regulating their canonical targets[328,331]. Therefore, dysregulation of isomiR processing could lead to a major shift in the miRNA pathway's activity that promotes oncogenesis. Future studies looking to investigate the role of miRNAs in leukemia and other cancers may benefit from incorporating isomiR-level comparisons.

Finally, like the miRNA pathway, alternative splicing is another form of gene regulation commonly perturbed in cancers[332,333]. However, it remains unknown how splicing of Argonaute may be affected

in many cancers and if misexpression of its isoforms contributes to miRNA dysregulation. Changes in Ago2 isoform expression may stabilize different miRNAs, which may in turn modulate the expression and activity of oncogenic or tumour suppressive miRNAs. Indeed, many of the miRNAs with the largest difference in abundance between Ago2 isoforms, including let-7b, miR-25, and miR-27b, have been described as oncomiRs[334–336].

Most studies in cancer research study the biological significance of miRNAs with the assumption that miRNA activity depends on expression of their canonical variants, and that wild-type Argonaute is the typical protein which forms the RISC[90,95,337]. However, our work suggests that two mechanisms – isomiR processing and Ago2 isoform expression – are key components of the miRNA pathway that can significantly impact miRNA activity in ways that don't always translate to changes in miRNA expression. This has serious implications that may explain, at least in part, why it has been so challenging to understand miRNA contributions to many cancers[338]. Furthermore, an incomplete understanding of the mechanisms that regulate the miRNA pathway may underlie some of the reasons that miRNA or siRNA-based therapies have mostly failed to translate into the clinic thus far[339,340]. It is clear from our work, and others, that there are significant shortcomings with our understanding of the miRNA pathway. However future research which builds on these findings may ultimately result in a greater chance of success with developing effective miRNA or siRNA-based therapies that improve patient outcomes.

5.3. miRNAs as biomarkers in cancer

Although studies have shown both RNA or miRNA expression profiles can classify cancers with a high degree of accuracy, several comparisons have shown improved performance with miRNAs and isomiRs over other classes of RNAs[200,341,342]. This suggests that miRNA and isomiR expression is tethered to cellular or tumour identity and argues for their utility as biomarkers in cancer. Their use as biomarkers is further encouraged by research suggesting many miRNAs are exported into the blood and remain stable for extended periods of time, creating the possibility of developing non-invasive diagnostic or prognostic tests that enable more personalized forms of therapy[146]. A recent study showed that miRNAs from the Dlk1-Dio3 locus were detectable from exosomes isolated from healthy patient blood samples, however we are not aware of any studies which have identified their presence in the blood of patients with glioblastoma[343]. Despite this, the fact that there is clearly a mechanism for their export into blood makes them promising biomarkers if they are indeed associated with clinically relevant cancer cell states.

5.4. Concluding Remarks

Despite limitations with the available data for single cell miRNA sequencing, we were able to find evidence of miRNA heterogeneity in glioblastoma. We identified two miRNA clusters that appear to be regulated by cell autonomous mechanisms and may be involved in cell state regulation. Additionally, we provide evidence that isomiRs are tightly controlled and can alter miRNA target regulation within individual cells, highlighting an additional layer of regulation that may contribute to intra-tumour heterogeneity. Finally, we showed that alternatively splicing can produce variants of Ago2 with altered functions through their selection of different miRNAs.

Appendix A - Non-Coding RNAs in Pediatric Solid Tumors

STATEMENT OF CONTRIBUTION OF OTHERS

Smith, C. M., Catchpoole D. & Hutvagner, G. Non-Coding RNAs in Pediatric Solid Tumors. *Frontiers in Genetics*. 10, 1-18 (2019).

I attest that the Higher Degree Research candidate Christopher Michael Smith was the primary contributor to the development of this publication.

This contribution included: planning the scope of the literature review and writing the manuscript.

Production Note:
Signature removed
prior to publication.

Daniel Catchpoole 06/05/2022

Production Note:
Signature removed
prior to publication.

Gyorgy Hutvagner 06/05/2022

frontiers in Genetics

REVIEW
published: 20 September 2019
doi: 10.3389/fgene.2019.00798

Non-Coding RNAs in Pediatric Solid Tumors

Christopher M. Smith[1], Daniel Catchpoole[2,3] and Gyorgy Hutvagner[1]**

[1] School of Biomedical Engineering, University of Technology Sydney, Sydney, Australia [2] School of Software, University of Technology Sydney, Sydney, Australia [3] The Tumour Bank–CCRU, Kids Research, The Children's Hospital at Westmead, Sydney, Australia

Pediatric solid tumors are a diverse group of extracranial solid tumors representing approximately 40% of childhood cancers. Pediatric solid tumors are believed to arise as a result of disruptions in the developmental process of precursor cells which lead them to accumulate cancerous phenotypes. In contrast to many adult tumors, pediatric tumors typically feature a low number of genetic mutations in protein-coding genes which could explain the emergence of these phenotypes. It is likely that oncogenesis occurs after a failure at many different levels of regulation. Non-coding RNAs (ncRNAs) comprise a group of functional RNA molecules that lack protein coding potential but are essential in the regulation and maintenance of many epigenetic and post-translational mechanisms. Indeed, research has accumulated a large body of evidence implicating many ncRNAs in the regulation of well-established oncogenic networks. In this review we cover a range of extracranial solid tumors which represent some of the rarer and enigmatic childhood cancers known. We focus on two major classes of ncRNAs, microRNAs and long non-coding RNAs, which are likely to play a key role in the development of these cancers and emphasize their functional contributions and molecular interactions during tumor formation.

Keywords: pediatric tumors, miRNA, long noncoding RNA, cancer biology, gene expression

OPEN ACCESS

Edited by:
Yun Zheng,
Kunming University of Science and Technology, China

Reviewed by:
Alessio Naccarati,
Italian Institute for Genomic Medicine (IIGM), Italy
Zexuan Zhu,
Shenzhen University, China

****Correspondence:***
Daniel Catchpoole
Daniel.Catchpoole@uts.edu.au
Gyorgy Hutvagner
gyorgy.hutvagner@uts.edu.au

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *24 January 2019*
Accepted: *30 July 2019*
Published: *20 September 2019*

Citation:
Smith CM, Catchpoole D and Hutvagner G (2019) Non-Coding RNAs in Pediatric Solid Tumors. Front. Genet. 10:798. doi: 10.3389/fgene.2019.00798

Pediatric cancers are often categorized as hematologic, intracranial, or extracranial (Chen et al., 2015). Hematologic cancers include those derived from the blood or blood forming tissues, including bone marrow and the lymph nodes. Intracranial cancers are tumors that develop inside the brain, whereas extracranial solid tumors, often referred to as pediatric solid tumors, arise outside the brain. Collectively, pediatric solid tumors represent approximately 40% of all pediatric cancers and commonly form in the developing sympathetic nervous system (neuroblastoma), retina (retinoblastoma), kidneys (Wilms tumor), liver (hepatoblastoma), bones (osteosarcoma, Ewing sarcoma), or muscles (rhabdomyosarcoma) (Kline and Sevier, 2003; Allen-Rhoades et al., 2018). Solid tumors can originate from cells of any of the three germ layers, the ectoderm, mesoderm, or endoderm, and likely arise due to disruptions in the developmental processes of these precursor cells, leading them to develop cancerous phenotypes (Chen et al., 2015). This contrasts with most adult cancers, which tend to be of epithelial origin and are believed to develop over time due to exposure to toxins and environmental stress. As a result, adult cancers often display a high occurrence of genetic mutations, whereas pediatric solid tumors tend to feature a relatively low number of genetic mutations. This has led to investigations into alternative forms of gene regulation that may contribute to the emergence and development of cancerous cells in pediatric cancers.

Non-coding RNAs (ncRNAs) form a group of functional RNAs lacking protein-coding potential, which play a crucial role in the regulation of gene expression at every level, from epigenetic regulation via methylation and chromatin packaging to post-transcriptional regulation (Cech and Steitz, 2014; Zhao et al., 2016). The most widely studied ncRNAs are the microRNAs (miRNAs), small 20- to 25-nucleotide-long RNAs that play an important role in regulating translation and messenger RNA stability via complementary base pairing (Huang et al., 2011). Other classes of small ncRNAs include small interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs), small transfer RNAs (tRFs), small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), and small cytoplasmic RNAs (scRNAs). Additionally, long non-coding RNAs (lncRNAs) are a loosely defined group of RNAs normally larger than 200 nucleotides long that lack protein-coding potential and do not fall into any of the other categories but nonetheless play key roles in the regulation of gene expression (Mercer et al., 2009). Increasing data suggest that ncRNAs play a role in regulating all biological processes, and it is no surprise that studies have observed widespread dysregulation of ncRNAs in nearly all forms of cancer (Prensner and Chinnaiyan, 2011; Leichter et al., 2017). Interestingly, dysregulated RNA patterns are often specific to the type of cancer or even subtype and can provide insight into the mechanisms underlying phenotypic differences between tumors or cells within a tumor, such as their aggressiveness or resistance to certain types of treatments (Blenkiron et al., 2007; Li et al., 2014). Additionally, genome-wide association studies have suggested that over 80% of single nucleotide polymorphisms found associated with cancer are outside of coding regions (Carninci et al., 2005; Cheetham et al., 2013). In this review, we will discuss how two major classes of ncRNAs, miRNAs and lncRNAs, may contribute to pediatric solid tumors by participating in the regulation of established oncogenic networks known to drive these cancers.

MICRORNAS AND GENE REGULATION

Not long after the first human miRNA, *let-7*, was discovered in 2002 by the Ruvkun lab, miRNAs began to emerge as key participants in tumorigenesis (Pasquinelli et al., 2000). In 2002, two miRNAs were identified as potential tumor suppressors due to their frequent downregulation or deletion in chronic lymphocytic leukemia (Calin et al., 2002). Calin et al. (2004) later showed that many miRNA genes are located close to fragile sites or common breakpoints that frequently occur in cancers, suggesting that their loss of function was a key event in oncogenesis. Since then, oncomiRs—cancer-associated miRNAs—have become a major research focus (Esquela-Kerscher and Slack, 2006). A better understanding of the mechanisms behind miRNA regulation in cancer is invaluable to researchers and clinicians alike, not only to aid in the identification of new drug targets but also for the development of promising RNA-based therapies and their potential use as early detection biomarkers.

miRNAs: Biogenesis and Functions

The life cycle of a miRNA typically begins with its transcription into a primary miRNA (pri-miRNA) by RNA polymerase II (Ha and Kim, 2014). pri-miRNAs share several similarities with messenger RNAs (mRNAs) in that they are 5' capped, are 3' polyadenylated, and can be several hundreds or thousands of nucleotides long. In many cases, the pri-miRNA encodes for one miRNA species; however, in humans, a substantial number are polycistronic and encode several different miRNAs together. pri-miRNAs must be processed in the nucleus by the RNAse III enzyme Drosha, which releases shorter ~65-nucleotide-long precursor RNAs (pre-miRNAs) with a secondary hairpin structure. This hairpin is recognized by the Exportin-5/Ran-GTP transporter, which transports the pre-miRNA from the nucleus to the cytoplasm. In the cytoplasm, the pre-miRNA is further processed by Dicer, another RNAse III enzyme, which cleaves the loop and releases a double-stranded miRNA duplex containing the 5 prime (5p) and 3 prime (3p) sequences. The duplex is then recognized by one of the four human Argonaute proteins, which loads one of the strands and discards the other.

miRNAs carry out their functions by binding to Argonaute and associating with various other proteins to form the RNA induced silencing complex (RISC). As part of this complex, miRNAs serve as guides by binding via complementary base pairing to target sites that are normally found in the 3′-untranslated region (3′UTR) of mRNAs. RISCs can regulate gene expression by direct cleavage of transcripts, transcript destabilization, or blocking translation. In a broader sense, miRNAs play a role in globally "fine-tuning" gene expression and are particularly important in inducing and maintaining differentiated cell states. In cancer, this finely tuned expression is often impaired, enabling gene networks that are normally switched on or off to reverse and begin influencing cellular behavior in a deleterious manner.

miRNAs: Drivers or Passengers in Cancer?

Microarrays and next-generation sequencing technologies enabled global measurements of miRNA expression changes and have revealed miRNA dysregulation to be a hallmark in nearly all cancers. miRNA expression profiles often correlate with cancer subtypes and have been effective at classifying cancer samples for risk stratification (De Preter et al., 2011). However, understanding the contribution of specific miRNAs can prove difficult. miRNAs are predicted to regulate hundreds to thousands of genes; however, their influence may be minor, and often, they must act in concert with other miRNAs. Current miRNA target prediction algorithms are imperfect and do not capture the true range of regulatory targets; therefore, biological validation is still needed (Riffo-Campos et al., 2016). Additionally, opposing behavior is seen with many miRNAs, where the same miRNA may be considered an oncogene in one cancer and a tumor suppressor in another. Because of their integration within complex gene networks, it is often not obvious whether a dysregulated miRNA actively participates in the maintenance of a cancerous state or whether it is simply a

bystander. Therefore, it is important to examine how miRNAs participate in oncogenic networks on a functional level in order to properly understand their role.

Transcription factors that play an important role in regulating cell proliferation, migration, and apoptosis are commonly perturbed in pediatric solid tumors. One of the best examples of this is in neuroblastoma, where *MYCN* amplification is present in approximately 25% of neuroblastoma patients and disproportionally represents high-risk cases (Huang and Weiss, 2013). MYCN upregulation is also observed at a higher frequency in several other pediatric solid tumors including Wilms tumor, rhabdomyosarcoma (Williamson et al., 2005), and retinoblastoma, although generally not to the extent seen in neuroblastoma. Germline inactivation of the Wilms Tumor 1 (WT1) transcription factor has been linked to a genetic predisposition towards Wilms tumor. Several transcription factors, including Twist, Snails, and Zebs, involved in the epithelial-to-mesenchymal transition have also been implicated in the development of osteosarcoma (Yang et al., 2013). miRNAs are often closely tied to transcription factors, either as regulators or as transcriptional targets (**Figure 1**) (Sin-Chan et al., 2019). One of the earliest studies linking miRNAs to an oncogenic transcription factor was by O'Donnell et al. in 2005 (O'Donnell et al., 2005). In this study, they demonstrated that c-Myc could induce expression of the miR-17~92 cluster and that several of these miRNAs could in turn regulate E2F1 transcription to control cell proliferation.

Disruptions in miRNA Processing

Recent studies have shown that impairments of the miRNA processing machinery are common in Wilms tumor and likely contribute to this disease. For example, a study by Torrezan et al. (2014) found mutations in miRNA processing genes in 33% of tumors, most commonly occurring in the *Drosha* gene, with other mutations in *DICER1*, *XPO5*, *DGCR8* and *TARBP2*. These results are supported by several other studies by Wu et al. (2013), Rakheja et al. (2014), Walz et al. (2015), Wegert et al. (2015), and Gadd et al. (2017). In Rakheja et al.'s study, they further examined the potential consequences of several of these mutations and found that Drosha mutations often led to a loss of RNAse IIIB activity, which prevented processing of

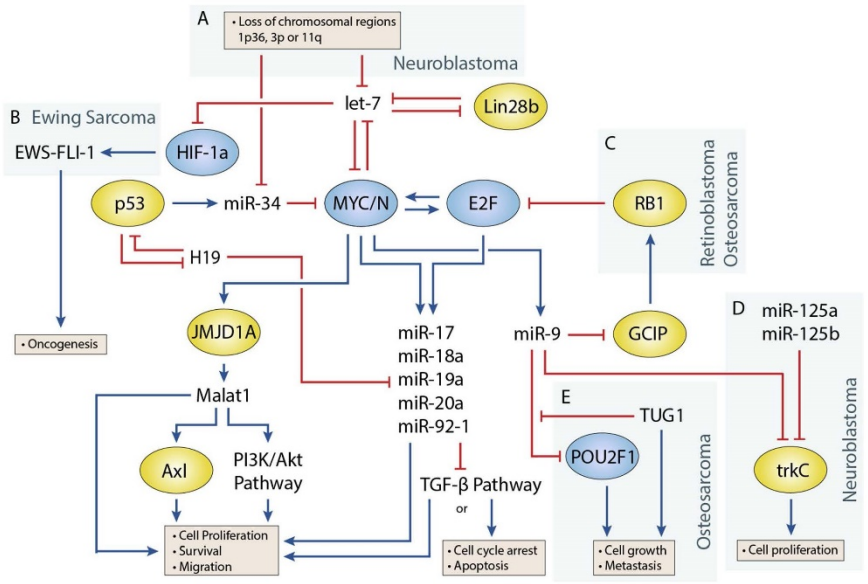

FIGURE 1 | Regulatory circuitry involving non-coding RNAs in various pediatric solid tumors. Shows elements of a key regulatory circuit involving MYC and E2F family transcription factors and many ncRNAs, often dysregulated in pediatric solid tumors. In many cases, recurring dysregulation of specific elements, including miRNAs and lncRNAs, is observed and may represent vulnerabilities in the normal development of specific cell lineages. **(A)** Loss of chromosomal regions where let-7 and miR-34 miRNAs are localized is frequently observed in neuroblastoma and may represent a key event in the development of many of these cancers. **(B)** let-7 dysregulation may facilitate overexpression of the oncogenic fusion transcript EWS-FLI-1 in Ewing sarcoma. **(C)** The RB1 tumor suppressor regulates E2F, and loss of function via mutations can lead to the development of retinoblastoma. In osteosarcoma, miR-9 may be able to act as an oncogenic driver as it is often overexpressed and can downregulate RB1. **(D)** In neuroblastoma, miR-9 can display tumor-suppressive properties by cooperating with miR-125a and miR-125b to regulate a specific isoform of trkC and suppress cell proliferation **(E)** The lncRNA TUG1 is suggested to act as a ceRNA against miR-9, which has been shown to display tumor-suppressive properties in some osteosarcoma cell lines.

pri-miRNAs, leading to a global reduction in mature miRNAs. *DICER1* mutations also frequently affected the RNAse IIIB domain; however, this mutation only affected processing of 5p miRNAs from precursors, as *DICER1* contains a second RNAse domain for 3p processing. As a result, this mutation led to a shift towards 3p miRNA maturation. These mutations have interesting consequences for global miRNA expression and most likely favor expression of oncogenic miRNAs or reduce expression of miRNAs with tumor-suppressive effects. In line with this, the let-7 family is predominantly 5p-derived, and lower expression of several of its 5p members was found in both *Drosha* and *DICER1* mutants in two of these studies (Rakheja et al., 2014; Walz et al., 2015). Additionally, the miR-200 family was found downregulated in Wilms tumors with mutated miRNA processing genes, which is known to regulate the mesenchymal-to-epithelial transition and has been associated with highly aggressive forms of cancer (Ceppi and Peter, 2014; Walz et al., 2015). The functional role of several oncomiRs has been investigated in detail within the context of pediatric solid tumors and is discussed in the following section.

The miR-17~92 Cluster is a Downstream Effector of Oncogenic Transcription Factors

The miR-17~92 cluster is expressed during normal development of the brain, heart, lungs, and immune system (Koralov et al., 2008; Ventura et al., 2008; Bian et al., 2013; Chen et al., 2013) and is known to regulate critical genes involved in cell growth, proliferation, and apoptosis. This cluster is comprised of six different miRNAs that are co-expressed, including miR-17, miR-18a, miR-19a, miR-19b-1, miR-20a, and miR-92-1. Dysregulation of the miR-17~92 cluster has been shown in several pediatric solid tumors including neuroblastoma, Wilms tumor, retinoblastoma, and osteosarcoma, where a higher expression generally correlates with a poorer prognosis (Chen and Stallings, 2007; Baumhoer et al., 2012; Li et al., 2014). The miR-17~92 cluster is particularly interesting due to its regulation by the transcription factor MYC and its homologue MYCN, where it seems to act as a mediator for some of MYC/MYCN's oncogenic effects (Schulte et al., 2008). Other transcription factors known to target the miR-17~92 cluster include members of the E2F family and STAT3 (Mogilyansky and Rigoutsos, 2013).

Several studies have demonstrated that the miR-17~92 cluster regulates many downstream components of the transforming growth factor beta (TGF-β) pathway, which is known to participate in a variety of cellular process such as differentiation, proliferation, and immune cell activation. A study by Fontana et al. (2008) demonstrated that in neuroblastoma, miR-17 and miR-20a downregulate the cyclin-dependent kinase inhibitor p21, which is activated by TGF-β. p21 plays a key role in the inhibition of cell cycle progression by blocking the transition from G1 to S phase, and its deregulation leads to uncontrolled cell growth. Additionally, Fontana et al. (2008) showed that miR-17-5p regulated another downstream component of TGF-β, the pro-apoptotic factor Bcl-2 interacting mediator (BIM). Mestdagh et al. (2010) later investigated miR-17~92 regulation of the TGF-β pathway in more depth and identified miR-17 and miR-20a as regulators of TGFBR2 and miR-18a as a regulator of SMAD2 and SMAD4, both signal transducers for TGF-β receptors. miR-18a and miR-19a have also been shown to repress estrogen receptor (ESR1) expression, and prolonged knockdown of miR-18a induced morphological differentiation of SK-N-BE neuroblastoma cells. Interestingly, the TGF-β pathway interacts with ESR1 signaling via several of the SMADs (Band and Laiho, 2011), suggesting a complex interplay between miR-17~92 and its targeted pathways necessary for fine-tuning differentiation during neuronal development—a balance that is disrupted when miR-17~92 is overexpressed. While no studies have investigated in detail the interaction between the miR-17~92 cluster and TGF-β pathway in Wilms tumor, the TGF-β pathway has been implicated in Wilms tumor development. In contrast with neuroblastoma, the TGF-β pathway appears to function as a promoter of Wilms tumor progression, and TGF-β is highly expressed in primary tumors, even more so in metastatic tumors. This multifaceted behavior of the TGF-β pathway has been shown in other cancers and implies that the pathway's influence is specific to the tumor it is activated in.

The E2F family of transcription factors serve an important role in cell cycle control as their expression can cause cells to enter the G1 phase to initiate cell division (Chen et al., 2009). Several members, including E2F1, E2F2, and E2F3, all regulate miR-17~92 expression. In a study by Kort et al. (2008) a member of the E2F family of transcription factors, E2F3, was shown to be exclusively expressed in Wilms tumor and not in other types of kidney tumors. In line with this, they compared expression of the miR-17~92 miRNAs in Wilms tumor samples to other renal tumor subtypes and found them all to be upregulated. They were also able to show a correlation between E2F3 expression and the stage of Wilms tumor, where it was highest in late-stage metastatic tissues. In retinoblastoma, an early study investigating the miR-17~92 cluster identified that one of its members, miR-20a, participates in an autoregulatory feedback loop with E2F2 and E2F3 (Sylvestre et al., 2007), as they found both transcription factors are themselves downregulated by miR-20a. The authors suggested that this autoregulation was critical in preventing expression of excessive amounts of E2F transcription factors. Given that MYC/MYCN and E2F have previously been shown to induce each other's expression, miR-20a appears to play an important role in keeping this positive feedback loop in check (Leone et al., 1997; Strieder and Lutz, 2003). Therefore, it is easy to see how disruption in one or more of these regulatory elements could lead to uncontrolled expression of these proliferative and anti-apoptotic signals.

A later study by Conkrite et al. (2011) investigated miR-17~92 in retinoblastoma and revealed that this cluster was capable of driving retinoblastoma formation in *RB1/p107*-deficient mice. *RB1* plays a key role in inhibiting cell cycle progression, and germline mutations of this gene can lead to familial retinoblastoma formation (Friend et al., 1986; Classon and Harlow, 2002). *RB1*'s protein product, pRB, inhibits E2F transcription factors by binding and inactivating them, and its absence enables miR-17~92–driven tumor formation.

The miR-17~92 cluster also plays a role in driving tumor progression and metastasis in osteosarcoma (Li et al., 2014). A recent study by Yang et al. (2018) identified QKI2 as a regulatory target of the miR-17~92 cluster. QKI proteins have previously

been shown to inhibit β-catenin and induce differentiation in colon cancer. Yang at el. demonstrated that miR-17~92 downregulated QKI2, causing upregulation of β-catenin, leading to increased proliferation, invasion, and migration in osteosarcoma (Yang et al., 2018). Additionally, miR-20a has previously been shown to downregulate Fas expression, which is a cell surface marker that interacts with FasL to induce apoptosis in the lungs, where osteosarcoma almost exclusively metastasizes to (Huang et al., 2012).

The miR-17~92 cluster plays a tumorigenic role in a number of pediatric solid tumors including neuroblastoma, Wilms tumor, retinoblastoma, and osteosarcoma. The use of the miRNA pathway by transcription factors such as the MYC and E2F families enables them to target a wide range of genes and immediately effect gene expression at the post-transcriptional level. Continued research into how miRNAs may operate as oncogenic drivers will likely expand the repertoire of potential drug targets available to us.

Let-7 Dysregulation is a Feature in Many Pediatric Solid Tumors

The let-7 family of miRNAs are among the most well-characterized tumor suppressors due to their frequent downregulation in cancers. In total, there are 12 members of the let-7 family located across eight different chromosomes; however, in most cells, only a selection of these miRNAs will be expressed (Balzeau et al., 2017). let-7 miRNAs are important in regulating the cell cycle and maintaining cells' differentiated state by targeting a wide range of genes with known roles in cancer biogenesis such as *MYC/MYCN*, *RAS*, *CDK6*, and *HMGA2* (Buechner et al., 2011; Wu et al., 2015).

let-7 is regulated by the LIN28 proteins, LIN28A and LIN28B, which mediate uridylation, prevent processing of the let-7 precursor, and are important for maintaining pluripotency in cells (Lehrbach et al., 2009; Balzeau et al., 2017). Both *Lin28* genes contain let-7 target sites and participate in a double-negative feedback loop with let-7 (Yin et al., 2017). Overexpression of *Lin28* tends to drive cells towards oncogenesis and is a common feature in cancers. In a study by Urbach et al. (2014), *Lin28b* overexpression was found in approximately 30% of Wilms tumors. Additionally, they found overexpression of *Lin28* could induce tumor formation in specific renal intermediates and that restoration of let-7 activity could reverse this effect in mice. Similar examples have been shown in mouse models, where *Lin28b* overexpression can drive hepatoblastoma and hepatocellular carcinoma in the liver and neuroblastoma in the neural crest (Molenaar et al., 2012; Nguyen et al., 2014). Molenaar et al. (2012) investigated Lin28b in neuroblastoma and demonstrated that Lin28b could enhance MYCN protein levels via let-7 regulation. However, a later study by Powers et al. (2016) showed that Lin28b expression was redundant in certain MYCN-amplified neuroblastoma cells, as overexpression of the MYCN transcript could function as a miRNA sponge for let-7, thereby negating their effect regardless of expression level. Powers et al. showed that most neuroblastomas were characterized by a loss of let-7 with either MYCN overexpression or chromosomal loss of arm 3p or 11q, where several let-7 miRNAs are located (**Figure 1A**). The authors noted that these events were generally mutually exclusive and suggested that the presence of one event alleviated selective pressure for the other.

A study by Di Fiore et al. (2016) revealed that let-7d could promote and suppress tumor formation within the same system. In this study, they found that let-7d overexpression in osteosarcoma cells reduced several stemness genes, including *Lin28b*, *HMGA2*, *Oct3/4*, and *SOX2*, and could elicit the mesenchymal-to-epithelial transition with upregulation of the epithelial marker E-cadherin and downregulation of mesenchymal markers N-cadherin and vimentin. However, they also found that let-7d enhanced cell migration and invasion, presumably by acting via the TGF-β pathway, which is known to promote this behavior. let-7d strongly increased versican VI expression, which has previously been shown to activate the TGF-β pathway in osteosarcoma (Li S. et al., 2014).

In Ewing sarcoma, Hameiri-Grossman et al. (2015) found that let-7 downregulated the *Ras* oncogene, as well as the transcription factor HIF-1a, to reduce EWS-FLI-1 expression (**Figure 1B**). EWS-FLI-1 is a hybrid transcript that results from a translocation event involving EWS and FLI1, and translocation events such as this are present in nearly all Ewing sarcoma cases and are believed to drive the disease (Delattre et al., 1994).

Loss of Let-7 plays a key role in many pediatric solid tumors as its loss enables expression of transcription factors and other genes that participate in oncogenesis. This has been emphasized in neuroblastoma, where it has been suggested that loss of let-7 function is an essential event in tumor development and positions the miRNA pathway as a central player in pediatric solid tumors.

miR-9 Has Been Shown to Play Oncogenic and Tumor-Suppressive Roles in Different Pediatric Tumors

miR-9 is a highly conserved miRNA involved in several different cellular processes including cell proliferation, differentiation, and migration. Early studies revealed miR-9 to be highly expressed in the brain and play a role both during development and in the adult brain; however, miR-9 has also been associated with many cancers outside the brain, acting as an oncogene or tumor suppressor (Coolen et al., 2013). Mir-9 is upregulated by MYC/MYCN and plays a role in promoting tumor growth and metastasis in several cancers including breast cancer, osteosarcoma, and rhabdomyosarcoma, where it is often overexpressed (Iorio et al., 2005; Luo et al., 2017) However, in other cancers such as neuroblastoma, miR-9's role is less clear, and studies have argued for oncogenic and tumor suppressor functions (Laneve et al., 2007; Zhi et al., 2014).

The role of miR-9 in osteosarcoma appears to be in promoting cell growth and metastasis (Zhu et al., 2015; Qi et al., 2016). In a study by Zhu et al. (2015), miR-9 knockdown suppressed cell growth and migration of osteosarcoma cells. They were also able to show that miR-9 downregulated RB1 via the Grap2 and cyclin D interacting protein (GCIP), thereby promoting E2F-mediated cell division (**Figure 1C**). Similar behavior has been observed in the alveolar subtype of rhabdomyosarcoma, where miR-9 contributes to increased cell proliferation and migration (Missiaglia et al., 2017). In this study by Missiaglia

et al. (2017), miR-9 was shown to be induced by the PAX3/FOXO1 fusion gene via MYCN, which is specific to this subtype of rhabdomyosarcoma.

In neuroblastoma, miR-9 expression has been shown to be both up- and downregulated in different studies. An early study by Laneve et al. (2007) showed that miR-9 was downregulated in 50% of primary neuroblastoma samples, and follow-up experiments demonstrated that miR-9 could act together with miR-125a and miR-125b to suppress cell proliferation by targeting a truncated isoform of the neurotrophin receptor tropomyosin-related kinase C (trkC) (**Figure 1D**). However, a later study by Ma et al. (2010) found miR-9 to be a target of MYCN and that miR-9 expression correlated with MYCN and metastatic status in neuroblastoma tumors. In this same study (albeit in breast cancer cells), Ma et al. also demonstrated that miR-9 suppressed E-cadherin to activate β-catenin and promote the epithelial-to-mesenchymal transition. Mir-9 is frequently involved in promoting cell migration; however, its absence has also been shown to produce different responses such as cell cycle arrest or apoptosis in neurons depending on their origin (Bonev et al., 2011). The contradictory behavior seen with studies of miR-9 highlight the diverse roles that individual miRNAs can play, and more comprehensive studies are needed to identify the relevant contextual influences on miRNA behavior.

miR-34 Is a Key Regulator of the Cell Cycle and Drug Resistance in Pediatric Solid Tumors

The miR-34 family has garnered significant interest since its members were discovered to be direct transcriptional targets of the tumor suppressor and transcription factor p53 (Hermeking, 2010). The miR-34 family consists of three miRNAs encoded by two genes, *mir-34a* and *mir-34b/c*. All three miRNAs play a key role in regulating apoptosis and the cell cycle by inducing G1 phase arrest. One of the more interesting facts about *miR-34a* and *miR-34b/c* is their genomic locations, which are located on chromosomes 1p36 and 11q23, respectively, regions that are frequently lost in pediatric solid tumors (Ruteshouser et al., 2005; Wittmann et al., 2007). In particular, loss of 1p36 occurs in 20–30% of neuroblastoma cases and correlates with MYCN amplification (Caron et al., 1993; Maris et al., 1995), whereas loss of 11q23 in occurs in approximately 40% of cases but almost never occurs with MYCN amplification (**Figure 1A**) (Guo et al., 1999; Attiyeh et al., 2005). miR-34 members are also regulators of the MYC family, as miR-34a is known to regulate MYCN and miR-34b and mir-34c to regulate c-MYC (Wei et al., 2008).

Studies on mir-34a expression have identified frequent downregulation in neuroblastoma, osteosarcoma, and hepatoblastoma (Jiao et al., 2016). miR-34a is itself considered a tumor suppressor due to its involvement in cell cycle arrest and apoptosis (De Antonellis et al., 2014). In neuroblastoma, Cole et al. (2008) investigated the growth-inhibitory effects of several miRNAs mapping to common chromosomal aberrations by overexpressing them in cell lines. In most cases, overexpression did not lead to a noticeable change in phenotype; however, miR-34a and miR-34c induced significant growth inhibition in cell lines with 1p36 deletion. Growth inhibition and suppression of metastasis by miR-34a have also been shown in osteosarcoma by several studies, where members of key proliferative signal transduction pathways such as c-Met, DUSP1, and Eag1 were identified as regulatory targets (Yan et al., 2012; Wu X. et al., 2013; Gang et al., 2017). The miR-34 family also targets several members of the Notch signaling pathway, which has been linked to both oncogenic and tumor-suppressive roles depending on the cellular context. In osteosarcoma, activation of the Notch pathway is known to contribute to tumor growth, and miR-34a–mediated downregulation of this pathway likely contributes to its tumor-suppressive role. However, in Ewing sarcoma, a recent study investigating miR-34b suggested that it could act as an oncogene, promoting proliferation, migration, and invasion through Notch1 repression (Lu Q. et al., 2018). Prior studies have shown correlations between high mir-34a expression and patient survival, which would indicate a tumor-suppressive role for mir-34a (Nakatani et al., 2012; Marino et al., 2014). It is unclear why miR-34a and miR-34b would display contrasting effects given their shared targets, and further investigation is needed.

Several studies by Pu et al. (2016) and Pu et al. (2017) have suggested that miR-34a may also play a role in promoting multidrug resistance in osteosarcoma. In these studies, they found that miR-34a-5p enhanced multidrug resistance through downregulation of the *CD117* and *AGTR1* genes *in vitro*. CD117 is often highly expressed in drug-resistant tumors and is commonly used as a marker for stemness (Adhikari et al., 2010). In contrast, Nakatani et al. found that miR-34a increased chemosensitivity in Ewing sarcoma (Nakatani et al., 2012).

Other miRNAs Involved in Multiple Pediatric Solid Tumors

A substantial number of other miRNAs have been discovered with functional implications in multiple pediatric solid tumors. One such miRNA is miR-125b, which typically exhibits tumor-suppressive properties in cancers such as neuroblastoma, osteosarcoma, and Ewing sarcoma, where it is commonly dysregulated (Laneve et al., 2007; Li J. et al., 2014; Xiao et al., 2019). Previously, it was mentioned that miR-125b participates in a network with miR-125a and miR-9, regulating expression of a truncated trkC isoform to control neuroblastoma growth and differentiation (Laneve et al., 2007; Le et al., 2009). In osteosarcoma, miR-125b was found to regulate STAT3 by downregulating MAP kinase kinase 7 (MKK7), which inactivates STAT3 via dephosphorylation (Xiao et al., 2019). Loss of miR-125b and consequent overexpression of MKK7 led to increased tumor formation and poorer prognosis. In Ewing sarcoma, miR-125b is involved in regulating the PI3K signaling pathway; could inhibit cell proliferation, migration, and invasion; and induce apoptosis through suppression of PIK3CD (Li J et al., 2014). Conversely, in retinoblastoma, miR-125b is overexpressed and has shown oncogenic properties by promoting cell proliferation and migration and inhibiting apoptosis (Bai et al., 2016). Conflicting behavior with miR-125b

has been observed in many other cancers, which suggests that its role is highly dependent on cell identity (Sun et al., 2013).

miR-124 has been widely reported to act as a tumor suppressor by inhibiting cell growth and metastasis and acts as a key mediator of differentiation in several pediatric solid tumors (Peng et al., 2014; Feng et al., 2015; Zhao et al., 2017). In neuroblastoma, miR-124a increased the proportion of differentiated cells possessing neurite outgrowths (Le et al., 2009). In retinoblastoma, miR-124 participates in a regulatory network with lncRNAs Malat1 and XIST, which both function as oncogenes by enhancing growth and metastasis through downregulation of miR-124 (Liu S. et al., 2017; Hu et al., 2018). miR-124 itself was shown to target STAT3 to inhibit cell proliferation, migration, and invasion (Liu S. et al., 2016). In Ewing sarcoma, miR-124 expression is suppressed, and expression was found to reduce growth and metastasis *via* downregulation of mesenchymal genes such as *SLUG* and cyclin D2 (*CCND2*) (Li et al., 2017). Finally, in osteosarcoma, retinoblastoma, and Ewing sarcoma, miR-143 has been found to be dysregulated (De Vito et al., 2012; Li S. et al., 2014; Wang et al., 2016; Sun et al., 2018). For example, Li et al. investigated miR-143 function in osteosarcoma and showed that miR-143 participated in the TGF-β pathway by targeting versican, and TGF-β could reduce miR-143 expression to promote cell migration and invasion (Li S. et al., 2014). FOS-like antigen 2 (FOSL2) was also identified as a miR-143 target, which enhanced cell proliferation, migration, and invasion in the absence of miR-143 (Sun et al., 2018). Additional miRNA studies have been listed in **Table 1**.

TABLE 1 | miRNAs that have been shown to exhibit oncogenic or tumor-suppressive effects through functional studies in various pediatric solid tumors.

Cancer	Oncogenic miRNAs/clusters	Comments
Neuroblastoma	mir-15 (Xin et al., 2013), mir-17~92 (Fontana et al., 2008; Loven et al., 2010; Mestdagh et al., 2010), miR-93 (Chakrabarti et al., 2012), miR-380 (Swarbrick et al., 2010), miR-558 (Shohet et al., 2011; Qu et al., 2015)	mir-17~92 dysregulation is common in MYCN-amplified neuroblastomas.
Osteosarcoma	let-7d (Di Fiore et al., 2016), miR-9 (Zhu et al., 2015; Qi et al., 2016), miR-17~92 (Huang et al., 2012; Li X. et al., 2014; Lu et al., 2018a; Yang et al., 2018b), miR-34a (Pu et al., 2016; Pu et al., 2017), miR-214 (Rehei et al., 2018)	
Retinoblastoma	miR-17~92 (Conkrite et al., 2011; Nittner et al., 2012; Jo et al., 2014), miR-125b (Bai et al., 2016)	Loss of RB1 function may enable mir-17~92–mediated oncogenicity.
Wilms tumor	miR-19b (Liu G.-L. et al., 2017), miR-483 (Veronese et al., 2010; Liu M. et al., 2013), miR-1180 (Jiang and Li, 2018)	
Hepatoblastoma	miR-492 (von Frowein et al., 2018)	
Ewing sarcoma	mir-17~92 (Schwentner et al., 2017), miR-20b (Kawano et al., 2017), miR-34b, b (Lu et al., 2018b) miR-130b (Satterfield et al., 2017)	EWS-FLI-1 may upregulate mir-17~92 and miR-34b.
Cancer	**Tumor-suppressive miRNAs/clusters**	**Comments**
Neuroblastoma	let-7 (Buechner et al., 2011; Molenaar et al., 2012; Henrchen et al., 2015; Powers et al., 2016), mir-7-1 (Chakrabarti et al., 2012), mir-9 (Laneve et al., 2007), mir-10 (Foley et al., 2011), miR-34a (Welch et al., 2007; Cole et al., 2008; Tivnan et al., 2011), miR-34c (Cole et al., 2008), miR-101 (Buechner et al., 2011), miR-124a (Le et al., 2009), miR-125 (Laneve et al., 2007; Le et al., 2009), miR-145 (Zhang et al., 2014), miR-184 (Chen and Stallings, 2007; Foley et al., 2010; Tivnan et al., 2010), miR-193b (Roth et al., 2018), miR-202 (Buechner et al., 2011), miR-203 (Zhao et al., 2015), miR-449 (Buechner et al., 2011), miR-542 (Bray et al., 2011), miR-584 (Xiang et al., 2015), miR-591 (Shohet et al., 2011)	let-7 and mir-34 are regulators of the MYCN oncogene.
Osteosarcoma	let-7d (Di Fiore et al., 2016), miR-1 (Novello et al., 2013), miR-34 (Yan et al., 2012; Wu et al., 2013b; Gang et al., 2017; Wen et al., 2017), miR-125b (Le et al., 2009; Xiao et al., 2019), miR-133b (Novello et al., 2013), miR-134 (Thayanithy et al., 2012), miR-138 (Zhu et al., 2016), miR-143 (Li et al., 2014c; Sun et al., 2018), miR-195 (Han et al., 2015), miR-223 (Dong et al., 2016), miR-363 (Wang K. et al., 2018), miR-369 (Thayanithy et al., 2012), miR-382 (Thayanithy et al., 2012), miR-451 (Yuan et al., 2015), miR-454 (Niu et al., 2015), miR-485 (Du et al., 2018), miR-544 (Thayanithy et al., 2012), miR-590 (Wang W.T. et al., 2018), miR-708 (Chen and Zhou, 2018), miR-2682 (Zhang et al., 2018b)	miR-34 suppresses tumor growth in osteosarcoma but may also contribute to drug resistance.
Retinoblastoma	miR-101 (Lei et al., 2014), miR-124 (Liu S. et al., 2016), miR-143 (Wang et al., 2016)	STAT3 is a target of miR-124.
Rhabdomyosarcoma	miR-1 (Rao et al., 2010; Li et al., 2012), mir-22 (Bersani et al., 2016), miR-29 (Li et al., 2012), miR-133a (Rao et al., 2010), miR-206 (Li et al., 2012; Missiaglia et al., 2010)	miR-1 appears to play a key role in differentiation of several sarcomas.
Wilms tumor	let-7 (Urbach et al., 2014), miR-16 (Chen W. et al., 2018), miR-34a (Chen W. et al., 2018), mir-92a (Zhu et al., 2018b), miR-613 (Wang et al., 2017a)	miR-92a was shown to act as a tumor suppressor unlike what is observed in other pediatric solid tumors.
Hepatoblastoma	miR-26 (Zhang et al., 2018d)	miR-26 was shown to repress LIN28B in hepatoblastoma.
Ewing sarcoma	let-7 (Hameiri-Grossman et al., 2015; Kawano et al., 2015), miR-16 (Kawano et al., 2015), miR-22 (Parrish et al., 2015), miR-29b (Kawano et al., 2015), miR-30a (Franzetti et al., 2013), miR-30d (Ye et al., 2018), miR-31 (Kamuth et al., 2014), miR-34a (Nakatani et al., 2012; Ventura et al., 2016), miR-124 (Li et al., 2017), miR-125b (Li et al., 2014b), miR-143 (De Vito et al., 2012), miR-145 (Riggi et al., 2010; De Vito et al., 2012), miR-185 (Zhang et al., 2018c), miR-193b (Moore et al., 2017), miR-708 (Robin et al., 2012)	Several miRNAs such as let-7 and miR-145 are downregulated by EWS-FLI-1.

miRNAs Regulate All Aspects of Tumorigenesis

Widespread dysregulation of miRNAs is observed in many pediatric solid tumors, and functional studies have demonstrated that many of these miRNAs can drive or repress oncogenic pathways responsible for cell proliferation, apoptosis, angiogenesis, metastasis, and drug resistance. Importantly, miRNAs such as let-7 and miR-34 play a vital role in pediatric solid tumors by regulating established oncogenic transcription factors such as the MYC and E2F families (Wei et al., 2008; Buechner et al., 2011). Other miRNAs, such as the miR-17~92 cluster and miR-9, serve as downstream effectors for these transcription factors, although their exact role in tumorigenesis seems to depend on the overall transcriptional landscape (Schulte et al., 2008; Ma et al., 2010). In some cases, viewing miRNAs as oncogenes or tumor suppressors likely represents an oversimplification of their role in cancer, and a better understanding of their participation in oncogenic networks will be needed to clarify their exact contributions.

LONG NON-CODING RNAS REGULATE ONCOGENIC PATHWAYS IN PEDIATRIC SOLID TUMORS

For a long time, it was believed that the human genome was mostly comprised of "junk" DNA, despite pervasive transcription of much of the genome outside of protein-coding genes and other known RNAs at the time (Prensner and Chinnaiyan, 2011). Originally thought of as transcriptional noise, lncRNAs have now emerged as functional regulators of nearly all essential cellular processes including growth, differentiation, cell state maintenance, apoptosis, splicing, and epigenetic regulation. The first lncRNA, H19, was discovered in 1990 where an RNA molecule was found spliced and polyadenylated in a manner typical of mRNAs; however, it lacked an open reading frame and was believed to function as an untranslated RNA molecule (Brannan et al., 1990).

Often, lncRNAs participate within protein complexes and can operate as scaffolds, guides, decoys, or allosteric regulators. Many lncRNAs function as epigenetic regulators by interacting with proteins involved in chromatin remodeling and DNA methylation. Frequently, these lncRNAs will be cis-acting and regulate the regions near their transcribed location; however, some are trans-acting. Other lncRNAs function as competing RNAs (ceRNAs), which contain miRNA binding sites in a similar manner to mRNAs in order to compete and reduce the activity of miRNAs.

Several studies have investigated lncRNA expression in pediatric tumors and have successfully identified unique expression profiles in different cancers and tumor subtypes (Mitra et al., 2012; Brunner et al., 2012; Dong et al., 2014; Sahu et al., 2018). For example, Dong et al. (2014) compared hepatoblastoma samples to normal liver tissue in patients and found 2,736 differentially expressed lncRNAs. A study by Pandey et al. (2014) found 24 lncRNAs that could distinguish low- and high-risk neuroblastoma tumors. In a more recent study, Sahu et al. (2018) identified 16 differentially expressed lncRNAs that could be used to predict event-free survival with greater accuracy than other commonly used clinical risk factors. Mechanistic studies into many of these lncRNAs have revealed that they frequently act as an additional layer of regulation within established oncogenic networks involving protein-coding genes and miRNAs. While the field of lncRNAs is still relatively young, many studies have emerged that suggest that lncRNAs are far more integrated into existing gene networks than what has previously been appreciated (**Figure 1**). In the following section, the roles of some of the better-characterized lncRNAs in pediatric solid tumors will be discussed.

Malat1 Is Induced by MYCN in Neuroblastoma and Competes With Many miRNAs

One of the earliest lncRNAs to be associated with disease was Malat1 (metastasis-associated lung adenocarcinoma transcript 1), which was shown to associate with metastatic tumors in non–small cell lung cancer patients (Ji et al., 2003). Malat1 is abundantly expressed and highly conserved across species, unlike many other lncRNAs, and displays remarkably diverse functions in cellular processes including alternative splicing, nuclear organization, and epigenetic modulation. Studies have suggested an important role for Malat1 in brain development, as it is highly expressed in neurons and its depletion has been shown to affect synapse and dendrite development (Bernard et al., 2010; Chen et al., 2016). However, its importance has been questioned as other studies have found that Malat1-KO mice are viable with no discernable change in phenotype (Nakagawa et al., 2012; Zhang et al., 2012).

In addition to lung cancer, Malat1 is known to contribute to metastasis in other common types of cancer including hepatocellular carcinoma and bladder cancer, with evidence that it acts through induction of the epithelial-to-mesenchymal transition (Ying et al., 2012; Li G. et al., 2014; Yang et al., 2017). The role of Malat1 in several pediatric cancers has also been explored in recent studies. In neuroblastoma, Tee et al. (2014) recently identified a regulatory network involving N-Myc, Malat1, and the histone demethylase JMJD1A. They found that N-Myc upregulated JMJD1A via direct binding of its promoter region and that JMJD1A could demethylate histone H3K9 near the promoter region of Malat1, leading to its upregulation. MYCN-mediated upregulation of Malat1 provides one mechanism in which its amplification can lead to increased metastasis in neuroblastoma patients. Another study by Bi et al. (2017) also demonstrated that Malat1 regulated Axl expression, a transmembrane receptor tyrosine kinase, which is known to activate pathways involved in cell proliferation, survival, and migration. In osteosarcoma, Dong et al. (2015) demonstrated that Malat1 was highly expressed and could activate the PI3K/Akt pathway to promote proliferation and invasion.

Malat1 is known to interact with many miRNAs implicated in cancer. In osteosarcoma, several studies have shown Malat1 can function as a ceRNA for different miRNAs (Wang et al., 2017b; Liu K. et al., 2017b; Sun and Qin, 2018). miR-140-5p is a tumor suppressor that downregulates HDAC4, a histone deacetylase that contributes to tumorigenesis, and competitive binding by Malat1 with miR-140-5p was shown to increase HDAC4 activity (Sun

and Qin, 2018). Malat1 was also shown to compete with miR-144-3p binding to ROCK1/ROCK2, promoting proliferation and metastasis (Wang et al., 2017b). In a similar manner, Liu K. et al. (2017) found that Malat1 could regulate cell growth through high-mobility group protein B1 (HMGB1) via ceRNA activity with miR-142-3p and miR-129-5p. Finally, in retinoblastoma, Malat1 downregulated miR-124 activity, leading to activation of the transcription factor SLUG, which is also targeted by miR-124 (Liu S. et al., 2017). SLUG has a known role in the epithelial-to-mesenchymal transition by suppressing E-cadherin via the Wnt/B-catenin pathway (Prasad et al., 2009).

In addition to interactions with miRNAs, Malat1 has also been shown to be processed directly by the Drosha–DGCR8 microprocessor complex through binding sites in the 5' end of the transcript (Maclas et al., 2012). lncRNAs such as Malat1 cooperate with the miRNA pathway and a number of transcription factors and epigenetic factors to form a complex network responsible for regulating tumorigenesis. The capacity for Malat1 to drive proliferation and metastasis in pediatric solid tumors suggests that dysregulation of any of these regulatory components can be sufficient for the development of cancer and highlights the value of further research into the relatively new field of lncRNAs.

H19: lncRNA Dysregulation via Loss of Imprinting may Contribute to Tumorigenesis

H19 is a paternally imprinted gene that is typically expressed exclusively from the maternal allele. Early reports suggested that H19 functioned as a tumor suppressor capable of inhibiting cell growth (Hao et al., 1993; Zhang et al., 1993; Casola et al., 1997; Fukuzawa et al., 1999). Studies in childhood solid tumors such as hepatoblastoma, Wilms tumor, and embryonic rhabdomyosarcoma supported this idea, as all three cancers often exhibited reduced H19 expression and had frequently lost the maternal 11p15 chromosomal region housing this gene (Fukuzawa et al., 1999). Other studies, in osteosarcoma and retinoblastoma, suggested an oncogenic role for H19, as its upregulation and loss of imprinting were commonly seen (Chan et al., 2014; Li L. et al., 2018). This observation was also seen in many other cancers including breast cancer (Lottin et al., 2002). Recently, the Hedgehog signaling pathway, a regulator of differentiation known to participate in cancer development and metastasis, was shown to induce H19 expression (Chan et al., 2014).

Understanding the exact function of H19 has proved difficult; however, it was known to sit downstream of the insulin growth factor 2 (IGF2) gene, a growth factor known to play a role in tumorigenesis. Early reports suggested interactions between IGF2 and H19, as loss of imprinting of either gene caused biallelic expression of the other gene (Ulaner et al., 2003). Ulaner et al. (2003) proposed a model for H19 and IGF2 involving a CCTF-binding site seated between the two genes, which could facilitate the blocking of IGF2 or transcription of H19 depending on its methylation status. However, this model suggested that H19 may simply serve as a marker for epigenetic disruptions and left open the question of what H19's actual function is.

More recent studies have demonstrated a role for H19 in epigenetic regulation. H19 binds to several epigenetic regulators including S-adenosylhomocysteine hydrolase (SAHH), methyl-CpG–binding domain protein 1 (MBD1), and enhancer of zeste homolog 2 (EZH2) (Raveh et al., 2015; Zhou et al., 2015). H19 was found to inhibit SAHH, which led to downregulation of DNMT3B-mediated methylation. MBD1 binds methylated DNA and recruits other proteins to mediate transcriptional repression or histone methylation, and H19 was shown to recruit this protein to several genes including IGF2 (Monnier et al., 2013). Finally, EZH2 is a histone methyltransferase that forms part of the Polycomb repressive complex 2 (PRC2) (Sauliere et al., 2006; Zhou et al., 2015).

H19 also plays a role in maintaining cells in an undifferentiated state by associating with the KH-type splicing regulatory protein (KSRP). When multipotent mesenchymal cells were induced, H19 was found to dissociate with KSRP to promote several of its functions including the decay of unstable mRNAs and increasing the expression of specific miRNAs involved in proliferation and differentiation though association with Drosha and Dicer.

H19's role in cancer has been emphasized by studies highlighting its relationship to the tumor suppressor p53. The H19 locus reciprocally regulates p53, as p53 suppresses H19 transcription and H19 can inactivate p53 by directly interacting with it (Yang et al., 2012). Notably, H19 also encodes for a miRNA in its first exon, miR-675, which suppresses p53 and several other targets including Rb, Igf1r, and several SMAD and cadherin genes. In the absence of functional p53, H19 was shown to promote tumor proliferation and survival under hypoxic conditions. Later studies in colorectal cancer showed that H19 could induce EMT by acting as a ceRNA (Liang et al., 2015). ceRNA function has recently been shown in a retinoblastoma study, targeting the mir-17~92 cluster (Zhang A. et al., 2018a). In this study, they found that H19 contained seven functional binding sites for mir-17~92 and was able to sponge mir-17~92 activity. This led to a de-repression of genes such as *p21* and STAT3 targets *BCL2*, *BCL2L1*, and *BIRC5*.

In a review by Raveh et al. (2015) it was proposed that H19 may behave differently in a manner that was dependent on the developmental stage of the cell, which could explain the evidence suggesting both oncogenic and tumor-suppressive roles. Here, the authors found that H19 functioned as a promoter of differentiation during the embryonic period and that absence of H19 at this stage could leave cells vulnerable to forming cancer, thereby seemingly acting as a tumor suppressor. However, in adult cells, where it is not normally expressed, H19 could function as an oncogene by promoting tumor survival and metastasis (Matouk et al., 2015).

TUG1 Regulates Transcription Factors Through Competition With miRNAs in Osteosarcoma

Recent studies have investigated the role of lncRNA TUG1 as a prognostic factor and ceRNA in osteosarcoma. Ma et al. (2016) identified a correlation between upregulation of TUG1 and poor

prognosis and metastasis, which was also evident in plasma, and suggested a potential use as a biomarker for patients with osteosarcoma. TUG1 is known to act through ceRNA activity against a number of miRNAs including miR-9, miR-132, miR-144, miR-153, miR-212, and miR-335 (Xie et al., 2016; Cao et al., 2017; Wang et al., 2017a; Li G. et al., 2018; Li H. et al., 2018). These miRNAs are known to regulate pathways involved in proliferation, cell cycle control, migration, and apoptosis. For example, TUG1 was shown to mediate de-repression of the transcription factor POU class 2 homeobox1 (POU2F1) via downregulation of mir-9 (**Figure 1E**) (Xie et al., 2016). POU2F1 itself participates in various cellular processes including growth, metabolism, stem cell identity, and metastasis (Vázquez-Arreguín and Tantin, 2016). In another example by Cao et al. (2017) they found that TUG1 also regulates migration and the epithelial-to-mesenchymal transition via ceRNA action on miRNA-144-3p. miR-144-3p is a regulator of EZH2, and upregulation of EZH2 induced cell migration through the Wnt/β-catenin pathway (Cao et al., 2017). Studies have also demonstrated direct interactions between TUG1 and the Polycomb repressor complex; however, to our knowledge, this has not been investigated in pediatric solid tumors (Yang et al., 2011).

Other Long Non-Coding RNAs in Pediatric Solid Tumors

In addition to those mentioned above, there are a number of other lncRNAs that have been identified as potential oncogenes or tumor suppressors involved in the pathogenesis of pediatric solid tumors (see **Table 2**) (Chen et al., 2017; Pandey et al., 2015).

For example, in osteosarcoma, lncRNAs HOTAIR, SNHG16, SNHG12, THOR, PACER, MFI2, and HOTTIP have all been shown to promote tumor or cell growth (Li et al., 2016; Qian et al., 2016; Ruan et al., 2016; Yin et al., 2016; Chen W. et al., 2018; Su et al., 2019; Wang et al., 2019). HOTAIR is known to play a role in chromatin regulation by acting as a scaffold for PRC2 and lysine-specific histone demethylase 1 (LSD1), and can also act as a ceRNA for miR-217 (Gupta et al., 2010; Tsai et al., 2010; Wang et al., 2019). SNHG16 has been shown to act as a ceRNA for several miRNAs including miR-205 and miR-340 (Zhu C. et al., 2018; Su et al., 2019). Additionally, several lncRNAs are downregulated in osteosarcoma with potential tumor-suppressive activity such as loc285194, MEG3, and TUSC7 (Pasic et al., 2010; Cong et al., 2016; Shi et al., 2018). loc285194 has been identified as a transcriptional target of p53 and can downregulate miR-211 (Liu Q. et al., 2013). In another study, increased p53

TABLE 2 | lncRNAs that play a role in pediatric solid tumors. OS, osteosarcoma; RB, retinoblastoma; NB, neuroblastoma; WT, Wilms tumor; HB, hepatoblastoma; RMS, rhabdomyosarcoma; ES, Ewing sarcoma.

Long Non-coding RNA	Cancer	Cellular Functions	References
Malat1	OS, RB, NB	Upregulates—proliferation, survival, migration, invasion.	(Tee et al., 2014; Dong et al., 2015; Bi et al., 2017; Liu K. et al., 2017; Liu S. et al., 2017; Wang K. et al., 2017; Sun and Qin, 2018)
H19	OS, RB, WT, HB, RMS	Upregulates—proliferation, survival. Regulates cell fate/differentiation.	(Zhang et al., 1993; Casola et al., 1997; Fukuzawa et al., 1999; Chan et al., 2014; Matouk et al., 2015; Raveh et al., 2015; Li L. et al., 2018)
TUG1	OS	Upregulates—proliferation, survival, migration.	(Ma et al., 2016; Xie et al., 2016; Cao et al., 2017; Wang S. et al., 2017; Li G. et al., 2018; Li H. et al., 2018)
HOTAIR	OS, RB	Upregulates—proliferation, survival, migration, invasion.	(Yang G. et al., 2018; Wang et al., 2019)
HOTTIP	OS	Upregulates—proliferation, resistance.	(Li et al., 2016)
SNHG12	OS	Upregulates—proliferation, migration.	(Ruan et al., 2016)
SNHG16	OS	Upregulates—proliferation, survival, migration, invasion.	(Su et al., 2019; Zhu et al., 2018a)
THOR	OS, RB	Upregulates—proliferation, migration.	(Chen K.S. et al., 2018; Shang, 2018)
PACER	OS	Upregulates—proliferation, invasion.	(Qian et al., 2016)
MFI2	OS	Upregulates—proliferation, survival, migration, invasion.	(Yin et al., 2016)
loc285194	OS	Downregulates—proliferation.	(Pasic et al., 2010)
TUSC7	OS	Downregulates—proliferation.	(Cong et al., 2016)
MEG3	OS, RB	Downregulates—proliferation, survival, invasion.	(Gao et al., 2017; Shi et al., 2018)
EWSAT1	OS, ES	Upregulates—proliferation, metastasis.	(Howarth et al., 2014; Sun et al., 2016)
XIST	RB	Upregulates—proliferation, survival.	(Hu et al., 2018)
DANCR	RB	Upregulates—proliferation, migration, invasion.	(Wang J. X. et al., 2018)
HOXA11-AS	RB	Upregulates—proliferation, survival.	(Han et al., 2019)
PANDAR	RB	Upregulates—survival.	(Sheng et al., 2018)
lncUSMYcN	NB	Upregulates—proliferation.	(Liu et al., 2014; Liu S. et al., 2016)
NBAT-1	NB	Downregulates—proliferation, invasion. Regulates cell fate/differentiation.	(Pandey et al., 2014)
CASC15-S	NB	Downregulates—proliferation, migration.	(Russell et al., 2015)
LINC00473	WT	Upregulates—proliferation, survival.	(Zhu et al., 2018b)
CRNDE	HB	Upregulates—proliferation, angiogenesis.	(Dong et al., 2017)
LINC01314	HB	Downregulates—proliferation, migration.	(Lv et al., 2018)

expression and a decrease in cell proliferation and invasion were observed when MEG3 was overexpressed (Shi et al., 2018). Furthermore, MEG3 was found to be downregulated by another lncRNA, EWSAT1, which had previously been shown to enhance cell proliferation and metastasis in both osteosarcoma and Ewing sarcoma (Howarth et al., 2014; Sun et al., 2016).

In retinoblastoma, HOTAIR, THOR, and MEG3 appear to have a similar influence as seen in osteosarcoma, where they also acted as oncogenes (HOTAIR and THOR) or tumor suppressors (MEG3) (Gao et al., 2017; Shang, 2018; Yang G. et al., 2018). In the study examining HOTAIR in retinoblastoma, HOTAIR was shown to be engaged in a reciprocal regulatory loop with miR-613 and promoted cell proliferation and activation of the EMT, potentially through upregulation of N-cadherin, vimentin, and α-SMA (Yang G. et al., 2018). Several lncRNAs have also been found acting as oncogenic ceRNAs including XIST, DANCR, and HOXA11-AS (Hu et al., 2018; Wang J. X. et al., 2018; Han et al., 2019). Finally, PANDAR is upregulated in retinoblastoma and may regulate cell proliferation and apoptosis via the Bcl-2/caspase-3 pathway (Sheng et al., 2018).

A number of studies have also suggested an important role for lncRNAs in neuroblastoma. For example, lncUSMYcN is an lncRNA that is frequently co-amplified alongside MYCN (Liu P. Y. et al., 2016). Liu et al. found that in neuroblastoma, lncUSMycN could upregulate MYCN through transcriptional activation of NCYM (a.k.a. MYCNOS), which codes for a protein that stabilizes MYCN (Suenaga et al., 2014). NCYM RNA has also been suggested to bind to the RNA-binding protein NonO, which is also known to upregulate MYCN expression (Liu et al., 2014; Liu P. Y. et al., 2016). Neuroblastoma associated transcript-1 (NBAT-1) is an epigenetic regulator that interacts with EZH2, and functions as a tumor suppressor due to its important role in neuronal differentiation (Pandey et al., 2014). Loss of NBAT-1 expression was found to increase cell proliferation and invasion (Pandey et al., 2014). Finally, an isoform of lncRNA CASC15, CASC15-S, was also implicated as a key element in neuronal differentiation, and low expression was associated with a poor outcome in patients (Russell et al., 2015).

In Wilms tumor, a study by Zhu et al. identified LINC00473 as an oncogenic lncRNA that is upregulated in unfavorable tumors (Zhu et al., 2018b). LINC00473 was shown to promote tumor growth and metastasis by acting as a ceRNA for the tumor suppressor miR-195 (Zhu et al., 2018b).

A study by Dong et al. identified 1757 upregulated and 979 downregulated lncRNAs comparing hepatoblastoma and normal tissues, suggesting that lncRNAs play a key role in this disease as well (Dong et al., 2014). The lncRNAs Colorectal Neoplasia Differentially Expressed (CRNDE) and LINC01314 have been investigated in more detail in hepatoblastoma (Dong et al., 2017; Lv et al., 2018). CRNDE is known to be frequently upregulated in hepatoblastoma, and knockdown of CRNDE activated the mTOR pathway and inhibited tumor growth and angiogenesis with a corresponding decrease in VEGFA and Ang-2 levels (Dong et al., 2017). LINC01314 was identified as a tumor suppressor, reducing proliferation and migration via downregulation of cell cycle proteins MCM7 and cyclin D1 (Lv et al., 2018).

CONCLUDING REMARKS

It is now clear that both miRNAs and lncRNAs form integral parts of the biological networks known to be impaired in pediatric solid tumors. miRNAs such as let-7 and mir-34 are key regulators of many pediatric oncogenes including *MYC*, *MYCN*, *RAS*, and *MET* (Johnson et al., 2005; Wei et al., 2008; Buechner et al., 2011; Yan et al., 2012). Additionally, ncRNAs such as the miR-17~92 cluster, mir-9, and Malat1 also serve as downstream effectors of MYC and MYCN (Schulte et al., 2008; Ma et al., 2010; Tee et al., 2014). Many more ncRNAs participate in these and other pathways to form a highly complex regulatory network essential for maintaining an optimal cell state (See **Tables 1** and **2**). ncRNA dysregulation offers an alternative mechanism to genetic mutations and DNA methylation whereby cell development and differentiation can be disturbed. Despite the relatively rare occurrence of mutations in pediatric solid tumors, copy number variations are common and often occur at regions of the genome that harbor ncRNAs with tumor-suppressive roles (Wei et al., 2008; Powers et al., 2016). Gene expression is often imprecise; however, miRNAs provide a layer of robustness, which helps ensure that biological networks respond appropriately to signals and remain functional despite an ever-increasing cellular disorder (Ebert and Sharp, 2012). lncRNAs, too, play a vital role in maintaining order by forming RNA–protein complexes and serving as ceRNA antagonists against miRNA-mediated repression, although much more work is needed in this field to fully comprehend their range of biological roles. Functional studies have revealed that dysregulation of ncRNAs is capable of driving progenitor cells towards oncogenesis. For example, this has been shown in retinoblastoma, where overexpression of the mir-17~92 cluster could drive tumor formation in *RB/p107*-deficient mice (Conkrite et al., 2011).

While genome-wide association studies have revealed that miRNA processing is frequently disrupted in Wilms tumor, this has not been shown to the same extent in other pediatric solid tumors. However, genetic mutations of protein-coding genes are only one way in which disruptions of miRNA processing can be revealed. Most miRNA studies ignore the fact that a high proportion of expressed miRNAs are isoforms (isomiRs). isomiRs originating from the same miRNA gene can possess a great deal of functional variability, with differences in target acquisition or turnover rate that can have a significant impact on overall gene regulation. Studies focusing on isomiR expression will provide an additional layer of resolution to our understanding of miRNA dysregulation.

Recent developments in single-cell technology have revealed heterogeneity in gene expression profiles among individual cells in many cancers such as glioblastoma and neuroblastoma (Patel et al., 2014; Boeva et al., 2017). Such studies suggest that many tumors comprise different cellular subtypes with unique phenotypes such as growth rate, drug resistance, and metastatic potential, which demand a new way of approaching cancer treatments. miRNA expression in pediatric solid tumors may also be heterogenous; however, limitations in single-cell technologies have left this avenue relatively unexplored, and further developments are needed.

So far, ncRNA research has played a key role in advancing our understanding of the mechanisms behind pediatric solid tumor development. Evidence supports an active role for ncRNAs in cancer that extends beyond mere passengers. However, continued research is needed to fully comprehend the molecular events leading to the development of cancer and unlock new possibilities for drug targets and biomarkers, which will ultimately lead to a better outcome for patients afflicted by these diseases.

AUTHOR CONTRIBUTIONS

CS collected the information and wrote the review. DC and GH provided guidelines, consulted, and edited the manuscript.

FUNDING

This work was supported by the Australian Research Council DP180100120 project grant.

REFERENCES

Adhikari, A. S., Agarwal, N., Wood, B. M., Porretta, C., Ruiz, B., Pochampally, R. R., et al. (2010). CD117 and Stro-1 identify osteosarcoma tumor-initiating cells associated with metastasis and drug resistance. *Cancer Res.* 70, 4602–4612. doi: 10.1158/0008-5472.CAN-09-3463

Allen-Rhoades, W., Whittle, S. B., and Rainusso, N. (2018). Pediatric solid tumors of infancy: an overview. *Pediatr. Rev.* 39, 57–67. doi: 10.1542/pir.2017-0057

Attiyeh, E. F., London, W. B., Mossé, Y. P., Wang, Q., Winter, C., Khazi, D., et al. (2005). Chromosome 1p and 11q deletions and outcome in neuroblastoma. *N. Engl. J. Med.* 353, 2243–2253. doi: 10.1056/NEJMoa052399

Bai, S., Tian, B., Li, A., Yao, Q., Zhang, G., and Li, F. (2016). MicroRNA-125b promotes tumor growth and suppresses apoptosis by targeting DRAM2 in retinoblastoma. *Eye (Lond.)* 30, 1630–1638. doi: 10.1038/eye.2016.189

Balzeau, J., Menezes, M. R., Cao, S., and Hagan, J. P. (2017). The LIN28/let-7 pathway in cancer. *Front. Genet.* 8, 1–16. doi: 10.3389/fgene.2017.00031

Band, A. M., and Laiho, M. (2011). Crosstalk of TGF-β and estrogen receptor signaling in breast cancer. *J. Mammary Gland Biol. Neoplasia* 16, 109–115. doi: 10.1007/s10911-011-9203-7

Baumhoer, D., Zillmer, S., Unger, K., Rosemann, M., Atkinson, M. J., Irmler, M., et al. (2012). MicroRNA profiling with correlation to gene expression revealed the oncogenic miR-17-92 cluster to be up-regulated in osteosarcoma. *Cancer Genet.* 205, 212–219. doi: 10.1016/j.cancergen.2012.03.001

Bernard, D., Prasanth, K. V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., et al. (2010). A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.* 29, 3082–3093. doi: 10.1038/emboj.2010.199

Bersani, F., Lingua, M. F., Morena, D., Foglizzo, V., Miretti, S., Lanzetti, L., et al. (2016). Deep sequencing reveals a novel miR-22 regulatory network with therapeutic potential in rhabdomyosarcoma. *Cancer Res.* 76, 6095–6106. doi: 10.1158/0008-5472.CAN-16-0709

Bi, S., Wang, C., Li, Y., Zhang, W., Zhang, J., Lv, Z., et al. (2017). LncRNA-MALAT1–mediated Axl promotes cell invasion and migration in human neuroblastoma. *Tumor Biol.* 39, 1–7. doi: 10.1177/1010428317699796

Bian, S., Hong, J., Li, Q., Schebelle, L., Pollock, A., Knauss, J. L., et al. (2013). MicroRNA Cluster miR-17-92 regulates neural stem cell expansion and transition to intermediate progenitors in the developing mouse neocortex. *Cell Rep.* 3, 1398–1406. doi: 10.1016/j.celrep.2013.03.037

Blenkiron, C., Goldstein, L. D., Thorne, N. P., Spiteri, I., Chin, S.-F., Dunning, M. J., et al. (2007). MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol.* 8, R214. doi: 10.1186/gb-2007-8-10-r214

Boeva, V., Louis-Brennetot, C., Peltier, A., Durand, S., Pierre-Eugène, C., Raynal, V., et al. (2017). Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* 49, 1408–1413. doi: 10.1038/ng.3921

Bonev, B., Pisco, A., and Papalopulu, N. (2011). MicroRNA-9 reveals regional diversity of neural progenitors along the anterior–posterior axis. *Dev. Cell.* 20, 19–32. doi: 10.1016/j.devcel.2010.11.018

Brannan, C. I., Dees, E. C., Ingram, R. S., and Tilghman, S. M. (1990). The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.* 10, 28–36. doi: 10.1128/MCB.10.1.28

Bray, I., Tivnan, A., Bryan, K., Foley, N. H., Watters, K. M., Tracey, L., et al. (2011). MicroRNA-542-5p as a novel tumor suppressor in neuroblastoma. *Cancer Lett.* 303, 56–64. doi: 10.1016/j.canlet.2011.01.016

Brunner, A. L., Beck, A. H., Edris, B., Sweeney, R. T., Zhu, S. X., Li, R., et al. (2012). Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. *Genome Biol.* 13, R75. doi: 10.1186/gb-2012-13-8-r75

Buechner, J., Tømte, E., Haug, B. H., Henriksen, J. R., Løkke, C., Flægstad, T., et al. (2011). Tumour-suppressor microRNAs let-7 and mir-101 target the proto-oncogene MYCN and inhibit cell proliferation in MYCN-amplified neuroblastoma. *Br. J. Cancer* 105, 296–303. doi: 10.1038/bjc.2011.220

Calin, G. A., Dumitru, C. D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., et al. (2002). Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U.S A.* 99, 15524–15529. doi: 10.1073/pnas.242606799

Calin, G. A., Sevignani, C., Dumitru, C. D., Hyslop, T., Noch, E., Yendamuri, S., et al. (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl. Acad. Sci.* 101, 2999–3004. doi: 10.1073/pnas.0307323101

Cao, J., Han, X., Qi, X., Jin, X., and Li, X. (2017). TUG1 promotes osteosarcoma tumorigenesis by upregulating EZH2 expression via MIR-144-3p. *Int. J. Oncol.* 51, 1115–1123. doi: 10.3892/ijo.2017.4110

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., et al. (2005). Molecular biology: the transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563. doi: 10.1126/science.1112014

Caron, H., van Sluis, P., van Hoeve, M., de Kraker, J., Bras, J., Slater, R., et al. (1993). Allelic loss of chromosome 1p36 in neuroblastoma is of preferential maternal origin and correlates with N-myc amplification. *Nat. Genet.* 4, 187–190. doi: 10.1038/ng0693-187

Casola, S., Pedone, P. V., Cavazzana, A. O., Basso, G., Luksch, R., D'Amore, E. S. G., et al. (1997). Expression and parental imprinting of the H19 gene in human rhabdomyosarcoma. *Oncogene* 14, 1503–1510. doi: 10.1038/sj.onc.1200956

Cech, T. R., and Steitz, J. A. (2014). The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* 157, 77–94. doi: 10.1016/j.cell.2014.03.008

Ceppi, P., and Peter, M. E. (2014). MicroRNAs regulate both epithelial-to-mesenchymal transition and cancer stem cells. *Oncogene* 33, 269–278. doi: 10.1038/onc.2013.55

Chakrabarti, M., Khandkar, M., Banik, N. L., and Ray, S. K. (2012). Alterations in expression of specific microRNAs by combination of 4-HPR and EGCG inhibited growth of human malignant neuroblastoma cells. *Brain Res.* 1454, 1–13. doi: 10.1016/j.brainres.2012.03.017

Chan, L. H., et al. (2014). Hedgehog signaling induces osteosarcoma development through Yap1 and H19 overexpression. *Oncogene* 33, 4857–4866. doi: 10.1038/onc.2013.433

Cheetham, S. W., Gruhl, F., Mattick, J. S., and Dinger, M. E. (2013). Long noncoding RNAs and the genetics of cancer. *Br. J. Cancer* 108, 2419–2425. doi: 10.1038/bjc.2013.233

Chen, Y., and Stallings, R. L. (2007). Differential patterns of microRNA expression in neuroblastoma are correlated with prognosis, differentiation, and apoptosis. *Cancer Res.* 67, 976–983. doi: 10.1158/0008-5472.CAN-06-3667

Chen, G., and Zhou, H. (2018). MiRNA-708/CUL4B axis contributes into cell proliferation and apoptosis of osteosarcoma. *Eur. Rev. Med. Pharmacol. Sci.* 22, 5452–5459. doi: 10.26355/eurrev_201809_15805

Chen, H. Z., Tsai, S. Y., and Leone, G. (2009). Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat. Rev. Cancer* 9, 785–797. doi: 10.1038/nrc2696

Chen, J., Huang, Z.-P., Seok, H. Y., Ding, J., Kataoka, M., Zhang, Z., et al. (2013). mir-17–92 cluster is required for and sufficient to induce cardiomyocyte

proliferation in postnatal and adult hearts. *Circ. Res.* 112, 1557–1566. doi: 10.1161/CIRCRESAHA.112.300658

Chen, X., Pappo, A., and Dyer, M. A. (2015). Pediatric solid tumor genomics and developmental pliancy. *Oncogene* 34, 5207–5215. doi: 10.1038/onc.2014.474

Chen, L., Feng, P., Zhu, X., He, S., Duan, J., and Zhou, D. (2016). Long non-coding RNA Malat1 promotes neurite outgrowth through activation of ERK/MAPK signalling pathway in N2a cells. *J. Cell. Mol. Med.* 20, 2102–2110. doi: 10.1111/jcmm.12904

Chen, R., Wang, G., Zheng, Y., Hua, Y., and Cai, Z. (2017). Long non-coding RNAs in osteosarcoma. *Oncotarget* 8, 20462–20475. doi: 10.18632/oncotarget.14726

Chen, K. S., Stroup, E. K., Budhipramono, A., Rakheja, D., Nichols-Vinueza, D., Xu, L., et al. (2018). Mutations in microRNA processing genes in Wilms tumors derepress the IGF2 regulator PLAG1. *Genes Dev.* 32, 996–1007. doi: 10.1101/gad.313783.118

Chen, W., Chen, M., Xu, Y., Chen, X., Zhou, P., Zhao, X., et al. (2018). Long non-coding RNA THOR promotes human osteosarcoma cell growth in vitro and in vivo. *Biochem. Biophys. Res. Commun.* 499, 913–919. doi: 10.1016/j.bbrc.2018.04.019

Classon, M., and Harlow, E. (2002). The retinoblastoma tumour suppressor in development and cancer. *Nat. Rev. Cancer* 2, 910–917. doi: 10.1038/nrc950

Cole, K. A., Attiyeh, E. F., Mosse, Y. P., Laquaglia, M. J., Diskin, S. J., Brodeur, G. M., et al. (2008). A Functional screen identifies miR-34a as a candidate neuroblastoma tumor suppressor gene. *Mol. Cancer Res.* 6, 735–742. doi: 10.1158/1541-7786.MCR-07-2102

Cong, M., Li, J., Jing, R., and Li, Z. (2016). Long non-coding RNA tumor suppressor candidate 7 functions as a tumor suppressor and inhibits proliferation in osteosarcoma. *Tumor Biol.* 37, 9441–9450. doi: 10.1007/s13277-015-4414-y

Conkrite, K., Sundby, M., Mukai, S., Michael Thomson, J., Mu, D., Hammond, S. M., et al. (2011). Mir-17~92 cooperates with RB pathway mutations to promote retinoblastoma. *Genes Dev.* 25, 1734–1745. doi: 10.1101/gad.17027411

Coolen, M., Katz, S., and Bally-Cuif, L. (2013). miR-9: a versatile regulator of neurogenesis. *Front. Cell. Neurosci.* 7, 1–11. doi: 10.3389/fncel.2013.00220

De Antonellis, P., Carotenuto, M., Vandenbussche, J., De Vita, G., Ferrucci, V., Medaglia, C., et al. (2014). Early targets of miR-34a in neuroblastoma. *Mol. Cell. Proteomics* 13, 2114–2131. doi: 10.1074/mcp.M113.035808

De Preter, K., Mestdagh, P., Vermeulen, J., Zeka, F., Naranjo, A., Bray, I., et al. (2011). miRNA expression profiling enables risk stratification in archived and fresh neuroblastoma tumor samples. *Clin. Cancer Res.* 17, 7684–7692. doi: 10.1158/1078-0432.CCR-11-0610

De Vito, C., Riggi, N., Cornaz, S., Suvà, M.-L., Baumer, K., Provero, P., et al. (2012). A TARBP2-dependent miRNA expression profile underlies cancer stem cell properties and provides candidate therapeutic reagents in Ewing sarcoma. *Cancer Cell* 21, 807–821. doi: 10.1016/j.ccr.2012.04.023

Delattre, O., Zucman, J., Melot, T., Garau, X. S., Zucker, J.-M., Lenoir, G. M., et al. (1994). The Ewing family of tumors — a subgroup of small-round-cell tumors defined by specific chimeric transcripts. *N. Engl. J. Med.* 331, 294–299. doi: 10.1056/NEJM199408043310503

Di Fiore, R., Drago-Ferrante, R., Pentimalli, F., Di Marzo, D., Forte, I. M., Carlisi, D., et al. (2016). Let-7d miRNA shows both antioncogenic and oncogenic functions in osteosarcoma-derived 3AB-OS cancer stem cells. *J. Cell. Physiol.* 231, 1832–1841. doi: 10.1002/jcp.25291

Dong, R., Jia, D., Xue, P., Cui, X., Li, K., Zheng, S., et al (2014). Genome-wide analysis of long noncoding RNA (lncRNA) expression in hepatoblastoma tissues. *PLoS One* 9, 1–9. doi: 10.1371/journal.pone.0085599

Dong, Y., Liang, G., Yuan, B., Yang, C., Gao, R., and Zhou, X. (2015). MALAT1 promotes the proliferation and metastasis of osteosarcoma cells by activating the PI3K/Akt pathway. *Tumour Biol.* 36, 1477–1486. doi: 10.1007/s13277-014-2631-4

Dong, J., Liu, Y., Liao, W., Liu, R., Shi, P., and Wang, L. (2016). miRNA-223 is a potential diagnostic and prognostic marker for osteosarcoma. *J. Bone Oncol.* 5, 74–79. doi: 10.1016/j.jbo.2016.05.001

Dong, R., Liu, X.-Q., Zhang, B.-B., Liu, B.-H., Zheng, S., and Dong, K.-R. (2017). Long non-coding RNA-CRNDE: a novel regulator of tumor growth and angiogenesis in hepatoblastoma. *Oncotarget* 8, 42087–42097. doi: 10.18632/oncotarget.14992

Du, K., Zhang, X., Lou, Z., Guo, P., Zhang, F., Wang, B., et al. (2018). MicroRNA485-3p negatively regulates the transcriptional co-repressor CtBP1 to control the oncogenic process in osteosarcoma cells. *Int. J. Biol. Sci.* 14, 1445–1456. doi: 10.7150/ijbs.26335

Ebert, M. S., and Sharp, P. A. (2012). Roles for MicroRNAs in conferring robustness to biological processes. *Cell* 149, 505–524. doi: 10.1016/j.cell.2012.04.005

Esquela-Kerscher, A., and Slack, F. J. (2006). Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer* 6, 259–269. doi: 10.1038/nrc1840

Feng, T., Xu, D., Tu, C., Li, W., Ning, Y., Ding, J., et al. (2015). miR-124 inhibits cell proliferation in breast cancer through downregulation of CDK4. *Tumor Biol.* 36, 5987–5997. doi: 10.1007/s13277-015-3275-8

Foley, N. H., Bray, I. M., Tivnan, A., Bryan, K., Murphy, D. M., Buckley, P. G., et al. (2010). MicroRNA-184 inhibits neuroblastoma cell survival through targeting the serine/threonine kinase AKT2. *Mol. Cancer* 9, 83. doi: 10.1186/1476-4598-9-83

Foley, N. H., Bray, I., Watters, K. M., Das, S., Bryan, K., Bernas, T., et al. (2011). MicroRNAs 10a and 10b are potent inducers of neuroblastoma cell differentiation through targeting of nuclear receptor corepressor 2. *Cell Death Differ.* 18, 1089–1098. doi: 10.1038/cdd.2010.172

Fontana, L., Fiori, M. E., Albini, S., Cifaldi, L., Giovinazzi, S., Forloni, M., et al. (2008). Antagomir-17-5p abolishes the growth of therapy-resistant neuroblastoma through p21 and BIM. *PLoS One* 3, e2236. doi: 10.1371/journal.pone.0002236

Franzetti, G.-A., Laud-Duval, K., Bellanger, D., Stern, M.-H., Sastre-Garau, X., and Delattre, O. (2013). MiR-30a-5p connects EWS-FLI1 and CD99, two major therapeutic targets in Ewing tumor. *Oncogene* 32, 3915–3921. doi: 10.1038/onc.2012.403

Friend, S. H., Bernards, R., Rogelj, S., Weinberg, R. A., Rapaport, J. M., Albert, D. M., et al. (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* 323, 643–646. doi: 10.1038/323643a0

Fukuzawa, R., Umezawa, A., Ochi, K., Urano, F., Ikeda, H., and Hata, J. I. (1999). High frequency of inactivation of the imprinted H19 gene in 'sporadic' hepatoblastoma. *Int. J. Cancer* 82, 490–497. doi: 10.1002/(SICI)1097-0215(19990812)82:4<490::AID-IJC4>3.0.CO;2-I

Gadd, S., Huff, V., Walz, A. L., Ooms, A. H.A.G., Armstrong, A. E., Gerhard, D. S., et al. (2017). A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nat. Genet.* 49, 1487–1494. doi: 10.1038/ng.3940

Gang, L., Qun, L., Liu, W.-D., Li, Y.-S., Xu, Y.-Z., and Yuan, D.-T. (2017). MicroRNA-34a promotes cell cycle arrest and apoptosis and suppresses cell adhesion by targeting DUSP1 in osteosarcoma. *Am. J. Transl. Res.* 9, 5388–5399.

Gao, Y., Huang, P., and Zhang, J. (2017). Hypermethylation of MEG3 promoter correlates with inactivation of MEG3 and poor prognosis in patients with retinoblastoma. *J. Transl. Med.* 15, 1–10. doi: 10.1186/s12967-017-1372-8

Guo, C., White, P. S., Weiss, M. J., Hogarty, M. D., Thompson, P. M., Stram, D. O., et al. (1999). Allelic deletion at 11q23 is common in MYCN single copy neuroblastomas. *Oncogene* 18, 4948–4957. doi: 10.1038/sj.onc.1202887

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076. doi: 10.1038/nature08975

Ha, M., and Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 509–524. doi: 10.1038/nrm3838

Hameiri-Grossman, M., Porat-Klein, A., Yaniv, I., Ash, S., Cohen, I. J., Kodman, Y., et al. (2015). The association between let-7, RAS and HIF-1α in Ewing sarcoma tumor growth. *Oncotarget* 6, 33834–33848. doi: 10.18632/oncotarget.5616

Han, K., Chen, X., Bian, N., Ma, B., Yang, T., Cai, C., et al. (2015). MicroRNA profiling identifies MiR-195 suppresses osteosarcoma cell metastasis by targeting CCND1. *Oncotarget* 6, 8875–8889. doi: 10.18632/oncotarget.3560

Han, N., Zuo, L., Chen, H., Zhang, C., He, P., and Yan, H. (2019). Long non-coding RNA homeobox A11 antisense RNA (HOXA11-AS) promotes retinoblastoma progression via sponging miR-506-3p. *Onco. Targets Ther.* 12, 3509–3517. doi: 10.2147/OTT.S195404

Hao, Y., Crenshaw, T., Moulton, T., Newcomb, E., and Tycko, B. (1993). Tumour-suppressor activity of H19 RNA. *Nature* 365, 764–767. doi: 10.1038/365764a0

Hennchen, M., Stubbusch, J., Abarchan-El Makhfi, I., Kramer, M., Deller, T., Pierre-Eugene, C., et al. (2015). Lin28B and Let-7 in the control of sympathetic neurogenesis and neuroblastoma development. *J. Neurosci.* 35, 16531–16544. doi: 10.1523/JNEUROSCI.2560-15.2015

Hermeking, H. (2010). The miR-34 family in cancer and apoptosis. *Cell Death Differ.* 17, 193–199. doi: 10.1038/cdd.2009.56

Howarth, M. M., Simpson, D., Ngok, S. P., Nieves, B., Chen, R., Siprashvili, Z., et al. (2014). Long noncoding RNA EWSAT1-mediated gene repression facilitates Ewing sarcoma oncogenesis. *J. Clin. Invest.* 124, 5275–5290. doi: 10.1172/JCI72124

Hu, C., Liu, S., Han, M., Wang, Y., and Xu, C. (2018). Knockdown of lncRNA XIST inhibits retinoblastoma progression by modulating the miR-124/STAT3 axis. *Biomed. Pharmacother.* 107, 547–554. doi: 10.1016/j.biopha.2018.08.020

Huang, M., and Weiss, W. A. (2013). Neuroblastoma and MYCN. *Cold Spring Harb. Perspect. Med.* 3, a014415. doi: 10.1101/cshperspect.a014415

Huang, Y., Shen, X. J., Zou, Q., Wang, S. P., Tang, S. M., and Zhang, G. Z. (2011). Biological functions of microRNAs: a review. *J. Physiol. Biochem.* 67, 129–139. doi: 10.1007/s13105-010-0050-6

Huang, G., Nishimoto, K., Zhou, Z., Hughes, D., and Kleinerman, E. S. (2012). miR-20a encoded by the miR-17-92 cluster increases the metastatic potential of osteosarcoma cells by regulating Fas expression. *Cancer Res.* 72, 908–916. doi: 10.1158/0008-5472.CAN-11-1460

Iorio, M. V., Ferracin, M., Liu, C. G., Veronese, A., Spizzo, R., Sabbioni, S., et al. (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* 65, 7065–7070. doi: 10.1158/0008-5472.CAN-05-1783

Ji, P., Diederichs, S., Wang, W., Böing, S., Metzger, R., Schneider, P. M., et al. (2003). MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 22, 8031–8041. doi: 10.1038/sj.onc.1206928

Jiang, X., and Li, H. (2018). MiR-1180-5p regulates apoptosis of Wilms' tumor by targeting p73. *Onco. Targets. Ther.* 11, 823–831. doi: 10.2147/OTT.S148684

Jiao, C., Zhu, A., Jiao, X., Ge, J., and Xu, X. (2016). Combined low miR-34s are associated with unfavorable prognosis in children with hepatoblastoma: a Chinese population-based study. *J. Pediatr. Surg.* 51, 1355–1361. doi: 10.1016/j.jpedsurg.2016.02.091

Jo, D. H., Kim, J. H., Cho, C. S., Cho, Y.-L., Jun, H. O., Yu, Y. S., et al. (2014). STAT3 inhibition suppresses proliferation of retinoblastoma through down-regulation of positive feedback loop of STAT3/miR-17-92 clusters. *Oncotarget* 5, 11513–11525. doi: 10.18632/oncotarget.2546

Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., et al. (2005). RAS is regulated by the let-7 microRNA family. *Cell* 120, 635–647. doi: 10.1016/j.cell.2005.01.014

Karnuth, B., Dedy, N., Spieker, T., Lawlor, E. R., Gattenlöhner, S., Ranft, A., et al. (2014). Differentially expressed miRNAs in Ewing sarcoma compared to mesenchymal stem cells: low miR-31 expression with effects on proliferation and invasion. *PLoS One* 9, e93067. doi: 10.1371/journal.pone.0093067

Kawano, M., Tanaka, K., Itonaga, I., Iwasaki, T., and Tsumura, H. (2015). c-Myc represses tumor-suppressive microRNAs, let-7a, miR-16 and miR-29b, and induces cyclin D2–mediated cell proliferation in Ewing's sarcoma cell line. *PLoS One* 10, e0138560. doi: 10.1371/journal.pone.0138560

Kawano, M., Tanaka, K., Itonaga, I., Iwasaki, T., and Tsumura, H. (2017). MicroRNA-20b promotes cell proliferation via targeting of TGF-β receptor II and upregulates MYC expression in Ewing's sarcoma cells. *Int. J. Oncol.* 51, 1842–1850. doi: 10.3892/ijo.2017.4155

Kline, N. E., and Sevier, N. (2003). Solid tumors in children. *J. Pediatr. Nurs.* 18, 96–102. doi: 10.1053/jpdn.2003.12

Koralov, S. B., Muljo, S. A., Galler, G. R., Krek, A., Chakraborty, T., Kanellopoulou, C., et al. (2008). Dicer ablation affects antibody diversity and cell survival in the B lymphocyte lineage. *Cell* 132, 860–874. doi: 10.1016/j.cell.2008.02.020

Kort, E. J., Farber, L., Tretiakova, M., Petillo, D., Furge, K. A., Yang, X. J., et al. (2008). The E2F3–oncomir-1 axis is activated in Wilms' tumor. *Cancer Res.* 68, 4034–4038. doi: 10.1158/0008-5472.CAN-08-0592

Laneve, P., Di Marcotullio, L., Gioia, U., Fiori, M. E., Ferretti, E., Gulino, A., et al. (2007). The interplay between microRNAs and the neurotrophin receptor tropomyosin-related kinase C controls proliferation of human neuroblastoma cells. *Proc. Natl. Acad. Sci.* 104, 7957–7962. doi: 10.1073/pnas.0700071104

Le, M. T. N., Xie, H., Zhou, B., Chia, P. H., Rizk, P., Um, M., et al. (2009). MicroRNA-125b promotes neuronal differentiation in human cells by repressing multiple targets. *Mol. Cell. Biol.* 29, 5290–5305. doi: 10.1128/MCB.01694-08

Lehrbach, N. J., Armisen, J., Lightfoot, H. L., Murfitt, K. J., Bugaut, A., Balasubramanian, S., et al. (2009). LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in *Caenorhabditis elegans*. *Nat. Struct. Mol. Biol.* 16, 1016–1020. doi: 10.1038/nsmb.1675

Lei, Q., Shen, F., Wu, J., Zhang, W., Wang, J., and Zhang, L. (2014). miR-101, downregulated in retinoblastoma, functions as a tumor suppressor in human retinoblastoma cells by targeting EZH2. *Oncol. Rep.* 32, 261–269. doi: 10.3892/or.2014.3167

Leichter, A. L., Sullivan, M. J., Eccles, M. R., and Chatterjee, A. (2017). MicroRNA expression patterns and signalling pathways in the development and progression of childhood solid tumours. *Mol. Cancer* 16, 1–17. doi: 10.1186/s12943-017-0584-0

Leone, G., DeGregori, J., Sears, R., Jakoi, L., and Nevins, J. R. (1997). Myc and Ras collaborate in inducing accumulation of active cyclin E/Cdk2 and E2F. *Nature* 387, 422–426. doi: 10.1038/387422a0

Li, L., Sarver, A. L., Alamgir, S., and Subramanian, S. (2012). Downregulation of microRNAs miR-1, -206 and -29 stabilizes PAX3 and CCND2 expression in rhabdomyosarcoma. *Lab. Investig.* 92, 571–583. doi: 10.1038/labinvest.2012.10

Li, G., Zhang, H., Wan, X., Yang, X., Zhu, C., Wang, A., et al. (2014). Long noncoding RNA plays a key role in metastasis and prognosis of hepatocellular carcinoma. *Biomed Res. Int.* 2014, 1–8. doi: 10.1155/2014/780521

Li, J., You, T., and Jing, J. (2014). MiR-125b inhibits cell biological progression of Ewing's sarcoma by suppressing the PI3K/Akt signalling pathway. *Cell Prolif.* 47, 152–160. doi: 10.1111/cpr.12093

Li, S., Li, F., and Cheng, T. (2014). TGF-β1 promotes osteosarcoma cell migration and invasion through the miR-143–versican pathway. *Cell. Physiol. Biochem.* 34, 2169–2179. doi: 10.1159/000369660

Li, X., Yang, H., Tian, Q., Liu, Y., and Weng, Y. (2014). Upregulation of microRNA-17-92 cluster associates with tumor progression and prognosis in osteosarcoma. *Neoplasma* 61, 453–460. doi: 10.4149/neo_2014_056

Li, Z., Zhao, L., and Wang, Q. (2016). Overexpression of long non-coding RNA HOTTIP increases chemoresistance of osteosarcoma cell by activating the Wnt/β-catenin pathway. *Am. J. Transl. Res.* 8, 2385–2393.

Li, Y., Shao, G., Zhang, M., Zhu, F., Zhao, B., He, C., et al. (2017). miR-124 represses the mesenchymal features and suppresses metastasis in Ewing sarcoma. *Oncotarget* 8, 10274–10286. doi: 10.18632/oncotarget.14394

Li, G., Liu, K., and Du, X. (2018). Long non-coding RNA TUG1 promotes proliferation and inhibits apoptosis of osteosarcoma cells by sponging miR-132-3p and upregulating SOX4 expression. *Yonsei Med. J.* 59, 226–235. doi: 10.3349/ymj.2018.59.2.226

Li, H., Tian, G., Tian, F., and Shao, L. (2018). Long non-coding RNA TUG1 promotes osteosarcoma cell proliferation and invasion through inhibition of microrna-212-3p expression. *Exp. Ther. Med.* 16, 779–787. doi: 10.3892/etm.2018.6216

Li, L., Chen, W., Wang, Y., Tang, L., and Han, M. (2018). Long non-coding RNA H19 regulates viability and metastasis, and is upregulated in retinoblastoma. *Oncol. Lett.* 8424–8432. doi: 10.3892/ol.2018.8385

Liang, W.-C., Fu, W.-M., Wong, C.-W., Wang, Y., Wang, W.-M., Hu, G.-X., et al. (2015). The lncRNA H19 promotes epithelial to mesenchymal transition by functioning as miRNA sponges in colorectal cancer. *Oncotarget* 6, 22513–22525. doi: 10.18632/oncotarget.4154

Liu, M., Roth, A., Yu, M., Morris, R., Bersani, F., Rivera, M. N., et al. (2013). The IGF2 intronic miR-483 selectively enhances transcription from IGF2 fetal promoters and enhances tumorigenesis. *Genes Dev.* 27, 2543–2548. doi: 10.1101/gad.224170.113

Liu, Q., Huang, J., Zhou, N., Zhang, Z., Zhang, A., Lu, Z., et al. (2013). LncRNA loc285194 is a p53-regulated tumor suppressor. *Nucleic Acids Res.* 41, 4976–4987. doi: 10.1093/nar/gkt182

Liu, P. Y., Erriquez, D., Marshall, G. M., Tee, A. E., Polly, P., Wong, M., et al. (2014). Effects of a novel long noncoding RNA, lncUSMycN, on N-Myc expression and neuroblastoma progression. *J. Natl. Cancer Inst.* 106, dju113–dju113. doi: 10.1093/jnci/dju113

Liu, P. Y., Atmadibrata, B., Mondal, S., Tee, A. E., and Liu, T. (2016). NCYM is upregulated by lncUSMycN and modulates N-Myc expression. *Int. J. Oncol.* 49, 2464–2470. doi: 10.3892/ijo.2016.3730

Liu, S., Hu, C., Wang, Y., Shi, G., Li, Y., and Wu, H. (2016). miR-124 inhibits proliferation and invasion of human retinoblastoma cells by targeting STAT3. *Oncol. Rep.* 36, 2398–2404. doi: 10.3892/or.2016.4999

Liu, G.-L., Yang, H.-J., Liu, B., and Liu, T. (2017). Effects of microRNA-19b on the proliferation, apoptosis, and migration of Wilms' tumor cells via the PTEN/

PI3K/AKT signaling pathway. *J. Cell. Biochem.* 118, 3424–3434. doi: 10.1002/jcb.25999

Liu, K., Huang, J., Ni, J., Song, D., Ding, M., Wang, J., et al. (2017). MALAT1 promotes osteosarcoma development by regulation of HMGB1 via miR-142–3p and miR-129–5p. *Cell Cycle* 16, 578–587. doi: 10.1080/15384101.2017.1288324

Liu, S., Yan, G., Zhang, J., and Yu, L. (2017). Knockdown of long noncoding RNA (lncRNA) metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) inhibits proliferation, migration, and invasion and promoted apoptosis by targeting miR-124 in retinoblastoma. *Oncol. Res. Featur. Preclin. Clin. Cancer Ther.* 26, 581–591. doi: 10.3727/096504017X14953948675403

Lottin, S., Adriaenssens, E., Dupressoir, T., Berteaux, N., Montpellier, C., Coll, J., et al. (2002). Overexpression of an ectopic H19 gene enhances the tumorigenic properties of breast cancer cells. Carcinogenesis 23, 1885–1895. doi: 10.1093/carcin/23.11.1885

Loven, J., Zinin, N., Wahlstrom, T., Muller, I., Brodin, P., Fredlund, E., et al. (2010). MYCN-regulated microRNAs repress estrogen receptor-α (ESR1) expression and neuronal differentiation in human neuroblastoma. *Proc. Natl. Acad. Sci.* 107, 1553–1558. doi: 10.1073/pnas.0913517107

Lu, C., Peng, K., Guo, H., Ren, X., Hu, S., Cai, Y., et al. (2018). miR-18a-5p promotes cell invasion and migration of osteosarcoma by directly targeting IRF2. *Oncol. Lett.* 16, 3150–3156. doi: 10.3892/ol.2018.9032

Lu, Q., Lu, M., Li, D., and Zhang, S. (2018). MicroRNA-34b promotes proliferation, migration and invasion of Ewing's sarcoma cells by downregulating Notch1. *Mol. Med. Rep.* 18, 3577–3588. doi: 10.3892/mmr.2018.9365

Luo, T., Yi, X., and Si, W. (2017). Identification of miRNA and genes involving in osteosarcoma by comprehensive analysis of microRNA and copy number variation data. *Oncol. Lett.* 14, 5427–5433. doi: 10.3892/ol.2017.6845

Lv, B., Zhang, L., Miao, R., Xiang, X., Dong, S., Lin, T., et al. (2018). Comprehensive analysis and experimental verification of LINC01314 as a tumor suppressor in hepatoblastoma. *Biomed. Pharmacother.* 98, 783–792. doi: 10.1016/j.biopha.2018.01.013

Ma, L., Young, J., Prabhala, H., Pan, E., Mestdagh, P., Muth, D., et al. (2010). MiR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nat. Cell Biol.* 12, 247–256. doi: 10.1038/ncb2024

Ma, B., Li, M., Zhang, L., Huang, M., Lei, J.-B., Fu, G.-H., Liu, C.-X., et al. (2016). Upregulation of long non-coding RNA TUG1 correlates with poor prognosis and disease status in osteosarcoma. *Tumor Biol.* 37, 4445–4455. doi: 10.1007/s13277-015-4301-6

MacIas, S., Plass, M., Stajuda, A., Michlewski, G., Eyras, E., and Cáceres, J. F. (2012). DGCR8 HITS-CLIP reveals novel functions for the microprocessor. *Nat. Struct. Mol. Biol.* 19, 760–766. doi: 10.1038/nsmb.2344

Marino, M. T., Grilli, A., Baricordi, C., Manara, M. C., Ventura, S., Pinca, R. S., et al. (2014). Prognostic significance of miR-34a in Ewing sarcoma is associated with cyclin D1 and ki-67 expression. *Ann. Oncol.* 25, 2080–2086. doi: 10.1093/annonc/mdu249

Maris, J. M., White, P. S., Beltinger, C. P., Sulman, E. P., Castleberry, R. P., Shuster, J. J., et al. (1995). Significance of chromosome 1p loss of heterozygosity in neuroblastoma. *Cancer Res.* 55, 4664–4669.

Matouk, I. J., Halle, D., Raveh, E., Gilon, M., Sorin, V., and Hochberg, A. (2015). The role of the oncofetal H19 lncRNA in tumor metastasis: orchestrating the EMT–MET decision. *Oncotarget* 7, 3748–3765. doi: 10.18632/oncotarget.6387

Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* 10, 155–159. doi: 10.1038/nrg2521

Mestdagh, P., Boström, A.-K., Impens, F., Fredlund, E., Van Peer, G., De Antonellis, P., et al. (2010). The miR-17-92 microRNA cluster regulates multiple components of the TGF-β pathway in neuroblastoma. *Mol. Cell* 40, 762–773. doi: 10.1016/j.molcel.2010.11.038

Missiaglia, E., Shepherd, C. J., Patel, S., Thway, K., Pierron, G., Pritchard-Jones, K., et al. (2010). MicroRNA-206 expression levels correlate with clinical behaviour of rhabdomyosarcomas. *Br. J. Cancer* 102, 1769–1777. doi: 10.1038/sj.bjc.6605684

Missiaglia, E., Shepherd, C. J., Aladowicz, E., Olmos, D., Selfe, J., Pierron, G., et al. (2017). MicroRNA and gene co-expression networks characterize biological and clinical behavior of rhabdomyosarcomas. *Cancer Lett.* 385, 251–260. doi: 10.1016/j.canlet.2016.10.011

Mitra, S. A., Mitra, A. P., and Triche, T. J. (2012). A central role for long non-coding RNA in cancer. *Front. Genet.* 3, 1–9. doi: 10.3389/fgene.2012.00017

Mogilyansky, E., and Rigoutsos, I. (2013). The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ.* 20, 1603–1614. doi: 10.1038/cdd.2013.125

Molenaar, J. J., Domingo-Fernández, R., Ebus, M. E., Lindner, S., Koster, J., Drabek, K., et al. (2012). LIN28B induces neuroblastoma and enhances MYCN levels via let-7 suppression. *Nat. Genet.* 44, 1199–1206. doi: 10.1038/ng.2436

Monnier, P., Martinet, C., Pontis, J., Stancheva, I., Ait-Si-Ali, S., and Dandolo, L. (2013). H19 lncRNA controls gene expression of the imprinted gene network by recruiting MBD1. *Proc. Natl. Acad. Sci.* 110, 20693–20698. doi: 10.1073/pnas.1310201110

Moore, C., Parrish, J. K., and Jedlicka, P. (2017). MiR-193b, downregulated in Ewing sarcoma, targets the ErbB4 oncogene to inhibit anchorage-independent growth. *PLoS One* 12, e0178028. doi: 10.1371/journal.pone.0178028

Nakagawa, S., Ip, J. Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T., et al. (2012). Malat1 is not an essential component of nuclear speckles in mice. *RNA* 18, 1487–1499. doi: 10.1261/rna.033217.112

Nakatani, F., Ferracin, M., Manara, M. C., Ventura, S., Del Monaco, V., Ferrari, S., et al. (2012). miR-34a predicts survival of Ewing's sarcoma patients and directly influences cell chemo-sensitivity and malignancy. *J. Pathol.* 226, 796–805. doi: 10.1002/path.3007

Nguyen, L. H., Robinton, D. A., Seligson, M. T., Wu, L., Li, L., Rakheja, D., et al. (2014). Lin28b is sufficient to drive liver cancer and necessary for its maintenance in murine models. *Cancer Cell* 26, 248–261. doi: 10.1016/j.ccr.2014.06.018

Nittner, D., Lambertz, I., Clermont, F., Mestdagh, P., Köhler, C., Nielsen, S. J., et al. (2012). Synthetic lethality between Rb, p53 and Dicer or miR-17-92 in retinal progenitors suppresses retinoblastoma formation. *Nat. Cell Biol.* 14, 958–965. doi: 10.1038/ncb2556

Niu, G., Li, B., Sun, J., and Sun, L. (2015). miR-454 is down-regulated in osteosarcomas and suppresses cell proliferation and invasion by directly targeting c-Met. *Cell Prolif.* 48, 348–355. doi: 10.1111/cpr.12187

Novello, C., Pazzaglia, L., Cingolani, C., Conti, A., Quattrini, I., Manara, M. C., et al. (2013). MiRNA expression profile in human osteosarcoma: role of miR-1 and miR-133b in proliferation and cell cycle control. *Int. J. Oncol.* 42, 667–675. doi: 10.3892/ijo.2012.1717

O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V., and Mendell, J. T. (2005). c-Myc–regulated microRNAs modulate E2F1 expression. *Nature* 435, 839–843. doi: 10.1038/nature03677

Pandey, G. K., Mitra, S., Subhash, S., Hertwig, F., Kanduri, M., Mishra, K., et al. (2014). The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. *Cancer Cell* 26, 722–737. doi: 10.1016/j.ccell.2014.09.014

Pandey, G. K., and Kanduri, C. (2015). Long noncoding RNAs and neuroblastoma. *Oncotarget* 6, 18265–18275. doi: 10.18632/oncotarget.4251

Parrish, J. K., Sechler, M., Winn, R. A., and Jedlicka, P. (2015). The histone demethylase KDM3A is a microRNA-22-regulated tumor promoter in Ewing sarcoma. *Oncogene* 34, 257–262. doi: 10.1038/onc.2013.541

Pasic, I., Shlien, A., Durbin, A. D., Stavropoulos, D. J., Baskin, B., Ray, P. N., et al. (2010). Recurrent focal copy-number changes and loss of heterozygosity implicate two noncoding RNAs and one tumor suppressor gene at chromosome 3q13.31 in osteosarcoma. *Cancer Res.* 70, 160–171. doi: 10.1158/0008-5472.CAN-09-1902

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408, 86–89. doi: 10.1038/35040556

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (80-.)* 344, 1396–1401. doi: 10.1126/science.1254257

Peng, X. H., Huang, H. R., Lu, J., Liu, X., Zhao, F. P., Zhang, B., et al. (2014). MiR-124 suppresses tumor growth and metastasis by targeting Foxq1 in nasopharyngeal carcinoma. *Mol. Cancer* 13, 1–13. doi: 10.1186/1476-4598-13-186

Powers, J. T., Tsanov, K. M., Pearson, D. S., Roels, F., Spina, C. S., Ebright, R., et al. (2016). Multiple mechanisms disrupt the let-7 microRNA family in neuroblastoma. *Nature* 535, 246–251. doi: 10.1038/nature18632

Prasad, C. P., Rath, G., Mathur, S., Bhatnagar, D., Parshad, R., and Ralhan, R. (2009). Expression analysis of E-cadherin, Slug and GSK3beta in invasive ductal carcinoma of breast. *BMC Cancer* 9, 325. doi: 10.1186/1471-2407-9-325

Prensner, J. R., and Chinnaiyan, A. M. (2011). The emergence of lncRNAs in cancer biology. *Cancer Discov.* 1, 391–407. doi: 10.1158/2159-8290.CD-11-0209

Pu, Y., Zhao, F., Wang, H., Cai, W., Gao, J., Li, Y., et al. (2016). MiR-34a-5p promotes the multi-drug resistance of osteosarcoma by targeting the CD117 gene. *Oncotarget* 7, 28420–28434. doi: 10.18632/oncotarget.8546

Pu, Y., Zhao, F., Li, Y., Cui, M., Wang, H., Meng, X., et al. (2017). The miR-34a-5p promotes the multi-chemoresistance of osteosarcoma via repression of the AGTR1 gene. *BMC Cancer* 17, 45. doi: 10.1186/s12885-016-3002-x

Qi, X. J., Wang, J. F., Wang, G. D., Xu, Q., and Sun, H. L. (2016). Pivotal role of microRNA-9 in osteosarcoma tumorigenesis and tumor progression. *Genet. Mol. Res.* 15, 1–10. doi: 10.4238/gmr.15017318

Qian, M., Yang, X., Li, Z., Jiang, C., Song, D., Yan, W., et al. (2016). P50-associated COX-2 extragenic RNA (PACER) overexpression promotes proliferation and metastasis of osteosarcoma cells by activating COX-2 gene. *Tumor Biol.* 37, 3879–3886. doi: 10.1007/s13277-015-3838-8

Qu, H., Zheng, L., Pu, J., Mei, H., Xiang, X., Zhao, X., et al. (2015). miRNA-558 promotes tumorigenesis and aggressiveness of neuroblastoma cells through activating the transcription of heparanase. *Hum. Mol. Genet.* 24, 2539–2551. doi: 10.1093/hmg/ddv018

Rakheja, D., Chen, K. S., Liu, Y., Shukla, A. A., Schmid, V., Chang, T. C., et al. (2014). Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. *Nat. Commun.* 2, 1–10. doi: 10.1038/ncomms5802

Rao, P. K., Missiaglia, E., Shields, L., Hyde, G., Yuan, B., Shepherd, C. J., et al. (2010). Distinct roles for miR-1 and miR-133a in the proliferation and differentiation of rhabdomyosarcoma cells. *FASEB J.* 24, 3427–3437. doi: 10.1096/fj.09-150698

Raveh, E., Matouk, I. J., Gilon, M., and Hochberg, A. (2015). The H19 long non-coding RNA in cancer initiation, progression and metastasis—a proposed unifying theory. *Mol. Cancer* 14, 1–14. doi: 10.1186/s12943-015-0458-2

Rehei, A.-L., Zhang, L., Pu, Y.-X., Mu, W.-B., Yang, D.-S., Liu, Y., et al. (2018). MicroRNA-214 functions as an oncogene in human osteosarcoma by targeting TRAF3. *Eur. Rev. Med. Pharmacol. Sci.* 22, 5156–5164. doi: 10.26355/eurrev_201808_15711

Riffo-Campos, Á. L., Riquelme, I., and Brebi-Mieville, P. (2016). Tools for sequence-based miRNA target prediction: what to choose? *Int. J. Mol. Sci.* 17, E1987. doi: 10.3390/ijms17121987

Riggi, N., Suvà, M.-L., De Vito, C., Provero, P., Stehle, J.-C., Baumer, K., et al. (2010). EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. *Genes Dev.* 24, 916–932. doi: 10.1101/gad.1899710

Robin, T. P., Smith, A., McKinsey, E., Reaves, L., Jedlicka, P., and Ford, H. L. (2012). EWS/FLI1 Regulates EYA3 in Ewing sarcoma via modulation of miRNA-708, resulting in increased cell survival and chemoresistance. *Mol. Cancer Res.* 10, 1098–1108. doi: 10.1158/1541-7786.MCR-12-0086

Roth, S. A., Hald, Ø. H., Fuchs, S., Løkke, C., Mikkola, I., Flægstad, T., et al. (2018). MicroRNA-193b-3p represses neuroblastoma cell growth via downregulation of Cyclin D1, MCL-1 and MYCN. *Oncotarget* 9, 18160–18179. doi: 10.18632/oncotarget.24793

Ruan, W., Wang, P., Feng, S., Xue, Y., and Li, Y. (2016). Long non-coding RNA small nucleolar RNA host gene 12 (SNHG12) promotes cell proliferation and migration by upregulating angiomotin gene expression in human osteosarcoma cells. *Tumor Biol.* 37, 4065–4073. doi: 10.1007/s13277-015-4256-7

Russell, M. R., Penikis, A., Oldridge, D. A., Alvarez-Dominguez, J. R., McDaniel, L., Diamond, M., et al. (2015). CASC15-S is a tumor suppressor lncRNA at the 6p22 neuroblastoma susceptibility locus. *Cancer Res.* 75, 3155–3166. doi: 10.1158/0008-5472.CAN-14-3613

Ruteshouser, E. C., Hendrickson, B. W., Colella, S., Krahe, R., Pinto, L., and Huff, V. (2005). Genome-wide loss of heterozygosity analysis of WT1-wild-type and WT1-mutant Wilms tumors. *Genes Chromosom. Cancer* 43, 172–180. doi: 10.1002/gcc.20169

Sahu, D., Ho, S.-Y., Juan, H.-F., and Huang, H.-C. (2018). High-risk, expression-based prognostic long noncoding RNA signature in neuroblastoma. *JNCI Cancer Spectr.* 2, 1–12. doi: 10.1093/jncics/pky015

Satterfield, L., Shuck, R., Kurenbekova, L., Allen-Rhoades, W., Edwards, D., Huang, S., et al. (2017). miR-130b directly targets ARHGAP1 to drive activation of a metastatic CDC42–PAK1–AP1 positive feedback loop in Ewing sarcoma. *Int. J. Cancer* 141, 2062–2075. doi: 10.1002/ijc.30909

Sauliere, J., Sureau, A., Expert-Bezancon, A., and Marie, J. (2006). The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. *Mol. Cell. Biol.* 26, 8755–8769. doi: 10.1128/MCB.00893-06

Schulte, J. H., Horn, S., Otto, T., Samans, B., Heukamp, L. C., Eilers, U. C., et al. (2008). MYCN regulates oncogenic microRNAs in neuroblastoma. *Int. J. Cancer* 122, 699–704. doi: 10.1002/ijc.23153

Schwentner, R., Herrero-Martin, D., Kauer, M. O., Mutz, C. N., Katschnig, A. M., Sienski, G., et al. (2017). The role of miR-17-92 in the miRegulatory landscape of Ewing sarcoma. *Oncotarget* 8, 10980–10993. doi: 10.18632/oncotarget.14091

Shang, Y. (2018). LncRNA THOR acts as a retinoblastoma promoter through enhancing the combination of c-myc mRNA and IGF2BP1 protein. *Biomed. Pharmacother.* 106, 1243–1249. doi: 10.1016/j.biopha.2018.07.052

Sheng, L., Wu, J., Gong, X., Dong, D., and Sun, X. (2018). SP1-induced upregulation of lncRNA PANDAR predicts adverse phenotypes in retinoblastoma and regulates cell growth and apoptosis in vitro and in vivo. *Gene* 668, 140–145. doi: 10.1016/j.gene.2018.05.065

Shi, Y., Lv, C., Shi, L., and Tu, G. (2018). MEG3 inhibits proliferation and invasion and promotes apoptosis of human osteosarcoma cells. *Oncol. Lett.* 15, 1917–1923. doi: 10.3892/ol.2017.7463

Shohet, J. M., Ghosh, R., Coarfa, C., Ludwig, A., Benham, A. L., Chen, Z., et al. (2011). A genome-wide search for promoters that respond to increased MYCN reveals both new oncogenic and tumor suppressor microRNAs associated with aggressive neuroblastoma. *Cancer Res.* 71, 3841–3851. doi: 10.1158/0008-5472.CAN-10-4391

Sin-Chan, P., Mumal, I., Suwal, T., Ho, B., Fan, X., Singh, I., et al. (2019). A C19MC–LIN28A–MYCN oncogenic circuit driven by hijacked super-enhancers is a distinct therapeutic vulnerability in ETMRs: a lethal brain tumor. *Cancer Cell.* 36, 51–67. doi: 10.1016/j.ccell.2019.06.002

Strieder, V., and Lutz, W. (2003). E2F proteins regulate MYCN expression in neuroblastomas. *J. Biol. Chem.* 278, 2983–2989. doi: 10.1074/jbc.M207596200

Su, P., Mu, S., and Wang, Z. (2019). Long noncoding RNA SNHG16 promotes osteosarcoma cells migration and invasion via sponging miRNA-340. *DNA Cell Biol.* 38, 170–175. doi: 10.1089/dna.2018.4424

Suenaga, Y., Islam, S. M., Alagu, J., Kaneko, Y., Kato, M., Tanaka, Y., et al. (2014). NCYM, a cis-antisense gene of MYCN, encodes a de novo evolved protein that inhibits GSK3β resulting in the stabilization of MYCN in human neuroblastomas. *PLoS Genet.* 10, e1003996. doi: 10.1371/journal.pgen.1003996

Sun, Y., and Qin, B. (2018). Long noncoding RNA MALAT1 regulates HDAC4-mediated proliferation and apoptosis via decoying of miR-140-5p in osteosarcoma cells. *Cancer Med.* 7, 4584–4597. doi: 10.1002/cam4.1677

Sun, Y. M., Lin, K. Y., and Chen, Y. Q. (2013). Diverse functions of miR-125 family in different cell contexts. *J. Hematol. Oncol.* 6, 1–8. doi: 10.1186/1756-8722-6-6

Sun, L., Yang, C., Xu, J., Feng, Y., Wang, L., and Cui, T. (2016). Long noncoding RNA EWSAT1 promotes osteosarcoma cell growth and metastasis through suppression of MEG3 expression. *DNA Cell Biol.* 35, 812–818. doi: 10.1089/dna.2016.3467

Sun, X., Dai, G., Yu, L., Hu, Q., Chen, J., and Guo, W. (2018). MiR-143-3p inhibits the proliferation, migration and invasion in osteosarcoma by targeting FOSL2. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-017-18739-3

Swarbrick, A., Woods, S. L., Shaw, A., Balakrishnan, A., Phua, Y., Nguyen, A., et al. (2010). miR-380-5p represses p53 to control cellular survival and is associated with poor outcome in MYCN-amplified neuroblastoma. *Nat. Med.* 16, 1134–1140. doi: 10.1038/nm.2227

Sylvestre, Y., De Guire, V., Querido, E., Mukhopadhyay, U. K., Bourdeau, V., Major, F., et al. (2007). An E2F/miR-20a autoregulatory feedback loop. *J. Biol. Chem.* 282, 2135–2143. doi: 10.1074/jbc.M608939200

Tee, A. E., Ling, D., Nelson, C., Atmadibrata, B., Dinger, M. E., Xu, N., et al. (2014). The histone demethylase JMJD1A induces cell migration and invasion by up-regulating the expression of the long noncoding RNA MALAT1. *Oncotarget* 5, 10000. doi: 10.18632/oncotarget.1785

Thayanithy, V., Sarver, A. L., Kartha, R. V., Li, L., Angstadt, A. Y., Breen, M., et al. (2012). Perturbation of 14q32 miRNAs-cMYC gene network in osteosarcoma. *Bone* 50, 171–181. doi: 10.1016/j.bone.2011.10.012

Tivnan, A., Foley, N. H., Tracey, L., Davidoff, A. M., and Stallings, R. L. (2010). MicroRNA-184-mediated inhibition of tumour growth in an orthotopic murine model of neuroblastoma. *Anticancer Res.* 30, 4391–4395.

Tivnan, A., Tracey, L., Buckley, P. G., Alcock, L. C., Davidoff, A. M., and Stallings, R. L. (2011). MicroRNA-34a is a potent tumor suppressor molecule in vivo in neuroblastoma. *BMC Cancer* 11, 33. doi: 10.1186/1471-2407-11-33

Torrezan, G. T., Ferreira, E. N., Nakahata, A. M., Barros, B. D. F., Castro, M. T. M., Correa, B. R., et al. (2014). Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour. *Nat. Commun.* 5, 4039. doi: 10.1038/ncomms5039

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., et al. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science (80-.)* 329, 689–693. doi: 10.1126/science.1192002

Ulaner, G. A., Vu, T. H., Li, T., Hu, J. F., Yao, X. M., Yang, Y., et al. (2003). Loss of imprinting of IGF2 and H19 in osteosarcoma is accompanied by reciprocal methylation changes of a CTCF-binding site. *Hum. Mol. Genet.* 12, 535–549. doi: 10.1093/hmg/ddg034

Urbach, A., Yermalovich, A., Zhang, J., Spina, C. S., Zhu, H., Perez-Atayde, A. R., et al. (2014). Lin28 sustains early renal progenitors and induces Wilms tumor. *Genes Dev.* 28, 971–982. doi: 10.1101/gad.237149.113

Vázquez-Arreguín, K., and Tantin, D. (2016). The Oct1 transcription factor and epithelial malignancies: old protein learns new tricks. *Biochim. Biophys. Acta* 1859, 792–804. doi: 10.1016/j.bbagrm.2016.02.007

Ventura, A., Young, A. G., Winslow, M. M., Lintault, L., Meissner, A., Erkeland, S. J., et al. (2008). Targeted deletion reveals essential and overlapping functions of the miR-17–92 family of miRNA clusters. *Cell* 132, 875–886. doi: 10.1016/j.cell.2008.02.019

Ventura, S., Aryee, D. N., Felicetti, F., De Feo, A., Mancarella, C., Manara, M. C., et al. (2016). CD99 regulates neural differentiation of Ewing sarcoma cells through miR-34a-Notch–mediated control of NF-κB signaling. *Oncogene* 35, 3944–3954. doi: 10.1038/onc.2015.463

Veronese, A., Lupini, L., Consiglio, J., Visone, R., Ferracin, M., Fornari, F., et al. (2010). Oncogenic role of miR-483-3p at the IGF2/483 locus. *Cancer Res.* 70, 3140–3149. doi: 10.1158/0008-5472.CAN-09-4456

von Frowein, J., Hauck, S. M., Kappler, R., Pagel, P., Fleischmann, K. K., Magg, T., et al. (2018). MiR-492 regulates metastatic properties of hepatoblastoma via CD44. *Liver Int.* 38, 1280–1291. doi: 10.1111/liv.13687

Walz, A. L., Ooms, A., Gadd, S., Gerhard, D. S., Smith, M. A., Guidry Auvil, J. M., et al. (2015). Recurrent DGCR8, DROSHA, and SIX homeodomain mutations in favorable histology Wilms tumors. *Cancer Cell* 27, 286–297. doi: 10.1016/j.ccell.2015.01.003

Wang, L.-L., Hu, H.-F., and Feng, Y.-Q. (2016). Suppressive effect of microRNA-143 in retinoblastoma. *Int. J. Ophthalmol.* 9, 1584–1590. doi: 10.18240/ijo.2016.11.08

Wang, H.-F., Zhang, Y.-Y., Zhuang, H.-W., and Xu, M. (2017). MicroRNA-613 attenuates the proliferation, migration and invasion of Wilms' tumor via targeting FRS2. *Eur. Rev. Med. Pharmacol. Sci.* 21, 3360–3369.

Wang, Y., Zhang, Y., Yang, T., Zhao, W., Wang, N., Li, P., et al. (2017a). Long non-coding RNA MALAT1 for promoting metastasis and proliferation by acting as a ceRNA of miR-144-3p in osteosarcoma cells. *Oncotarget* 8, 59417–59434. doi: 10.18632/oncotarget.19727

Wang, Y., Yang, T., Zhang, Z., Lu, M., Zhao, W., Zeng, X., et al. (2017b). Long non-coding RNA TUG1 promotes migration and invasion by acting as a ceRNA of miR-335-5p in osteosarcoma cells. *Cancer Sci.* 108, 859–867. doi: 10.1111/cas.13201

Wang, J. X., Yang, Y., and Li, K. (2018). Long noncoding RNA DANCR aggravates retinoblastoma through miR-34c and miR-613 by targeting MMP-9. *J. Cell. Physiol.* 233, 6986–6995. doi: 10.1002/jcp.26621

Wang, K., Yan, L., and Lu, F. (2018). MiR-363-3p inhibits osteosarcoma cell proliferation and invasion via targeting SOX4. *Oncol. Res.* 5, 157–163. doi: 10.3727/096504018X15190861873459

Wang, W.-T., Qi, Q., Zhao, P., Li, C.-Y., Yin, X.-Y., and Yan, R.-B. (2018). miR-590-3p is a novel microRNA which suppresses osteosarcoma progression by targeting SOX9. *Biomed. Pharmacother.* 107, 1763–1769. doi: 10.1016/j.biopha.2018.06.124

Wang, B., Qu, X. L., Liu, J., Lu, J., and Zhou, Z. Y. (2019). HOTAIR promotes osteosarcoma development by sponging miR-217 and targeting ZEB1. *J. Cell. Physiol.* 234, 6173–6181. doi: 10.1002/jcp.27394

Wegert, J., Ishaque, N., Vardapour, R., Geörg, C., Gu, Z., Bieg, M., et al. (2015). Mutations in the SIX1/2 Pathway and the DROSHA/DGCR8 miRNA microprocessor complex underlie high-risk blastemal type Wilms tumors. *Cancer Cell* 27, 298–311. doi: 10.1016/j.ccell.2015.01.002

Wei, J. S., Song, Y. K., Durinck, S., Chen, Q. R., Cheuk, A. T. C., Tsang, P., et al. (2008). The MYCN oncogene is a direct target of miR-34a. *Oncogene* 27, 5204–5213. doi: 10.1038/onc.2008.154

Welch, C., Chen, Y., and Stallings, R. L. (2007). MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. *Oncogene* 26, 5017–5022. doi: 10.1038/sj.onc.1210293

Wen, J., Zhao, Y. K., Liu, Y., and Zhao, J. F. (2017). MicroRNA-34a inhibits tumor invasion and metastasis in osteosarcoma partly by effecting C-IAP2 and Bcl-2. *Tumor Biol.* 39, 101042831770567. doi: 10.1177/1010428317705761

Williamson, D., Lu, Y. J., Gordon, T., Sciot, R., Kelsey, A., Fisher, C., et al. (2005). Relationship between MYCN copy number and expression in rhabdomyosarcomas and correlation with adverse prognosis in the alveolar subtype. *J. Clin. Oncol.* 23, 880–888. doi: 10.1200/JCO.2005.11.078

Wittmann, S., Zirn, B., Alkassar, M., Ambros, P., Graf, N., and Gessler, M. (2007). Loss of 11q and 16q in Wilms tumors is associated with anaplasia, tumor recurrence, and poor prognosis. *Genes Chromosom. Cancer* 46, 163–170. doi: 10.1002/gcc.20397

Wu, M. K., Sabbaghian, N., Xu, B., Addidou-Kalucki, S., Bernard, C., Zou, D., et al. (2013). Biallelic DICER1 mutations occur in Wilms tumours. *J. Pathol.* 230, 154–164. doi: 10.1002/path.4196

Wu, X., Zhong, D., Gao, Q., Zhai, W., Ding, Z., and Wu, J. (2013). MicroRNA-34a inhibits human osteosarcoma proliferation by downregulating ether à go-go 1 expression. *Int. J. Med. Sci.* 10, 676–682. doi: 10.7150/ijms.5528

Wu, L., Nguyen, L. H., Zhou, K., De Soysa, T. Y., Li, L., Miller, J. B., et al. (2015). Precise let-7 expression levels balance organ regeneration against tumor suppression. *Elife* 4, 1–16. doi: 10.7554/eLife.09431

Xiang, X., Mei, H., Qu, H., Zhao, X., Li, D., Song, H., et al. (2015). miRNA-584-5p exerts tumor suppressive functions in human neuroblastoma through repressing transcription of matrix metalloproteinase 14. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1852, 1743–1754. doi: 10.1016/j.bbadis.2015.06.002

Xiao, T., Zhou, Y., Li, H., Xiong, L., Wang, J., Wang, Z.-H., et al. (2019). MiR-125b suppresses the carcinogenesis of osteosarcoma cells via the MAPK–STAT3 pathway. *J. Cell. Biochem.* 120, 2616–2626. doi: 10.1002/jcb.27568

Xie, C.-H., Cao, Y.-M., Huang, Y., Shi, Q.-W., Guo, J.-H., et al. (2016). Long non-coding RNA TUG1 contributes to tumorigenesis of human osteosarcoma by sponging miR-9-5p and regulating POU2F1 expression. *Tumor Biol.* 37, 15031–15041. doi: 10.1007/s13277-016-5391-5

Xin, C., Buhe, B., Hongting, L., Chuanmin, Y., Xiwei, H., Hong, Z., et al. (2013). MicroRNA-15a promotes neuroblastoma migration by targeting reversion-inducing cysteine-rich protein with Kazal motifs (RECK) and regulating matrix metalloproteinase-9 expression. *FEBS J.* 280, 855–866. doi: 10.1111/febs.12074

Yan, K., Gao, J., Yang, T., Ma, Q., Qiu, X., Fan, Q., et al. (2012). MicroRNA-34a inhibits the proliferation and metastasis of osteosarcoma cells both in vitro and in vivo. *PLoS One.* 7, e33778. doi: 10.1371/journal.pone.0033778

Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K. A., Grinstein, J. D, et al. (2011). ncRNA- and Pc2 methylation–dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* 147, 773–788. doi: 10.1016/j.cell.2011.08.054

Yang, F., Bi, J., Xue, X., Zheng, L., Zhi, K., Hua, J., et al. (2012). Up-regulated long non-coding RNA H19 contributes to proliferation of gastric cancer cells. *FEBS J.* 279, 3159–3165. doi: 10.1111/j.1742-4658.2012.08694.x

Yang, G., Yuan, J., and Li, K. (2013). EMT transcription factors: implication in osteosarcoma. *Med. Oncol.* 30, 1–5. doi: 10.1007/s12032-013-0697-2

Yang, L., Li, Y., Wei, Z., and Chang, X. (2017). Coexpression network analysis identifies transcriptional modules associated with genomic alterations in neuroblastoma. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1864, 2341–2348. doi: 10.1016/j.bbadis.2017.12.020

Yang, G., Fu, Y., Lu, X., Wang, M., Dong, H., and Li, Q. (2018). LncRNA HOTAIR/miR-613/c-met axis modulated epithelial–mesenchymal transition of retinoblastoma cells. *J. Cell. Mol. Med.* 22, 5083–5096. doi: 10.1111/jcmm.13796

Yang, H., Peng, Z., Liang, M., Zhang, Y., Wang, Y., Jiang, Y., et al. (2018). The miR-17-92 cluster/QKI2/β-catenin osteosarcoma progression axis promotes 9, 25285–25293. doi: 10.18632/oncotarget.23935

Ye, C., Yu, X., Liu, X., Dai, M., and Zhang, B. (2018). miR-30d inhibits cell biological progression of Ewing's sarcoma by suppressing the MEK/ERK and PI3K/Akt pathways in vitro. *Oncol. Lett.* 15, 4390–4396. doi: 10.3892/ol.2018.7900

Yin, Z., Ding, H., He, E., Chen, J., and Li, M. (2016). Overexpression of long non-coding RNA MFI2 promotes cell proliferation and suppresses apoptosis in human osteosarcoma. *Oncol. Rep.* 36, 2033–2040. doi: 10.3892/or.2016.5013

Yin, J., Zhao, J., Hu, W., Yang, G., Yu, H., Wang, R., et al. (2017). Disturbance of the let-7/LIN28 double negative feedback loop is associated with radio- and chemo-resistance in non–small cell lung cancer. *PLoS One* 12, 1–16. doi: 10.1371/journal.pone.0172787

Ying, L., Chen, Q., Wang, Y., Zhou, Z., Huang, Y., and Qiu, F. (2012). Upregulated MALAT-1 contributes to bladder cancer cell migration by inducing epithelial-to-mesenchymal transition. *Mol. Biosyst.* 8, 2289–2294. doi: 10.1039/c2mb25070e

Yuan, J., Lang, J., Liu, C., Zhou, K., Chen, L., and Liu, Y. (2015). The expression and function of miRNA-451 in osteosarcoma. *Med. Oncol.* 32, 324. doi: 10.1007/s12032-014-0324-x

Zhang, Y., Shields, T., Crenshaw, T., Hao, Y., Moulton, T., and Tycko, B. (1993). Imprinting of human H19: allele-specific CpG methylation, loss of the active allele in Wilms tumor, and potential for somatic allele switching. *Am. J. Hum. Genet.* 53, 113–124.

Zhang, B., Arun, G., Mao, Y. S., Lazar, Z., Hung, G., Bhattacharjee, G., et al. (2012). The lncRNA malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep.* 2, 111–123. doi: 10.1016/j.celrep.2012.06.003

Zhang, H., Pu, J., Qi, T., Qi, M., Yang, C., Li, S., et al. (2014). MicroRNA-145 inhibits the growth, invasion, metastasis and angiogenesis of neuroblastoma cells through targeting hypoxia-inducible factor 2 alpha. *Oncogene* 33, 387–397. doi: 10.1038/onc.2012.574

Zhang, A., Shang, W., Nie, Q., Li, T., and Li, S. (2018). Long non-coding RNA H19 suppresses retinoblastoma progression via counteracting miR-17-92 cluster. *J. Cell. Biochem.* 119, 3497–3509. doi: 10.1002/jcb.26521

Zhang, F., Zhu, Y., Fan, G., and Hu, S. (2018). MicroRNA-2682-3p inhibits osteosarcoma cell proliferation by targeting CCND2, MMP8 and Myd88. *Oncol. Lett.* 16, 3359–3364. doi: 10.3892/ol.2018.9029

Zhang, S., Li, D., Jiao, G.-J., Wang, H.-L., and Yan, T.-B. (2018). miR-185 suppresses progression of Ewing's sarcoma via inhibiting the PI3K/AKT and Wnt/β-catenin pathways. *Onco. Targets. Ther.* 11, 7967–7977. doi: 10.2147/OTT.S167771

Zhang, Y., Zhao, Y., Wu, J., Liangpunsakul, S., Niu, J., and Wang, L. (2018). MicroRNA-26-5p functions as a new inhibitor of hepatoblastoma by repressing lin-28 homolog B and aurora kinase a expression. *Hepatol. Commun.* 2, 861–871. doi: 10.1002/hep4.1185

Zhao, D., Tian, Y., Li, P., Wang, L., Xiao, A., Zhang, M., et al. (2015). MicroRNA-203 inhibits the malignant progression of neuroblastoma by targeting Sam68. *Mol. Med. Rep.* 12, 5554–5560. doi: 10.3892/mmr.2015.4013

Zhao, Y., Sun, H., and Wang, H. (2016). Long noncoding RNAs in DNA methylation: new players stepping into the old game. *Cell Biosci.* 6, 1–6. doi: 10.1186/s13578-016-0109-3

Zhao, Y., Ling, Z., Hao, Y., Pang, X., Han, X., Califano, J. A., et al. (2017). MiR-124 acts as a tumor suppressor by inhibiting the expression of sphingosine kinase 1 and its downstream signaling in head and neck squamous cell carcinoma. *Oncotarget* 8, 25005–25020. doi: 10.18632/oncotarget.15334

Zhi, F., Wang, R., Wang, Q., Xue, L., Deng, D., Wang, S., et al. (2014). MicroRNAs in neuroblastoma: small-sized players with a large impact. *Neurochem. Res.* 39, 613–623. doi: 10.1007/s11064-014-1247-9

Zhou, J., Yang, L., Zhong, T., Mueller, M., Men, Y., Zhang, N., et al. (2015). H19 lncRNA alters DNA methylation genome wide by regulating S-adenosylhomocysteine hydrolase. *Nat. Commun.* 6, 1–13. doi: 10.1038/ncomms10221

Zhu, S. W., Li, J. P., Ma, X. L., Ma, J. X., Yang, Y., Chen, Y., et al. (2015). miR-9 modulates osteosarcoma cell growth by targeting the GCIP tumor suppressor. *Asian Pacific J. Cancer Prev.* 16, 4509–4513. doi: 10.7314/APJCP.2015.16.11.4509

Zhu, Z., Tang, J., Wang, J., Duan, G., Zhou, L., and Zhou, X. (2016). MIR-138 acts as a tumor suppressor by targeting EZH2 and enhances cisplatin-induced apoptosis in osteosarcoma cells. *PLoS One* 11, 1–12. doi: 10.1371/journal.pone.0150026

Zhu, C., Cheng, D., Qiu, X., Zhuang, M., and Liu, Z. (2018). Long noncoding RNA SNHG16 promotes cell proliferation by sponging microRNA-205 and upregulating ZEB1 expression in osteosarcoma. *Cell. Physiol. Biochem.* 51, 429–440. doi: 10.1159/000495239

Zhu, S., Fu, W., Zhang, L., Fu, K., Hu, J., Jia, W., et al. (2018a). LINC00473 antagonizes the tumour suppressor miR-195 to mediate the pathogenesis of Wilms tumour via IKKα. *Cell Prolif.* 51, 1–10. doi: 10.1111/cpr.12416

Zhu, S., Zhang, L., Zhao, Z., Fu, W., Fu, K., Liu, G., et al. (2018b). MicroRNA-92a-3p inhibits the cell proliferation, migration and invasion of Wilms tumor by targeting NOTCH1. *Oncol. Rep.* 40, 571–578. doi: 10.3892/or.2018.6458

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix B – Statistical Values and Read Counts for IsomiR-Target Pairs

miRNA	Category	KS Statistic	P.Value	Signif.	Mean Prop. of Total Reads	Delta Mean	Mean Read Count
miR-92a-3p	Total	0.0981	0.0219	*	0.7633	-0.057	19364.37
	Canonical	0.1334	0.000473	***	0.0173	-0.0841	453.3158
	5′ Variant	0.121	0.00209	**	0.6986	-0.0436	17701.95
	3′ Templated	0.1455	9.77E-05	***	0.183	-0.0709	4596.263
	3′ Non-templated (A)	0.2106	1.86E-09	***	0.4637	0.1122	11748.74
	3′ Non-templated (U)	0.1644	6.24E-06	***	0.1592	0.057	4107.526
	Substitutions	0.0864	0.0601		0.0509	-0.0333	1292.053
miR-146b-5p	Total	0.0367	0.984		0.04	0.0016	1046.474
	Canonical	0.1071	0.0555		0.0059	-0.0293	158.1579
	5′ Variant	0.1846	4.70E-05	***	0.0023	-0.0747	61
	3′ Templated	0.0398	0.965		0.0285	0.0102	744.1579
	3′ Non-templated (A)	0.0561	0.708		0.0026	-0.0073	68.7895
	3′ Non-templated (U)	0.0892	0.167		0.0009	-0.0501	24.8947
	Substitutions	0.1203	0.0217	*	0.0044	-0.0431	112.6316
miR-10a-5p	Total	0.0839	0.335		0.0303	0.0108	765.1579
	Canonical	0.0839	0.336		0.0023	-0.0419	61.2105
	5′ Variant	0.0766	0.447		0.0116	-0.0101	294.6842
	3′ Templated	0.0734	0.503		0.0233	0.027	576.8421
	3′ Non-templated (A)	0.0548	0.842		0.0019	-0.0077	50.0526
	3′ Non-templated (U)	0.1789	0.000606	**	0.0001	-0.085	2.8421
	Substitutions	0.1294	0.0289	*	0.0043	-0.0544	109.6316
miR-191-5p	Total	0.1074	0.367		0.0232	0.0116	576.1579
	Canonical	0.1275	0.185		0.0036	0.017	87.7368
	5′ Variant	0.1687	0.031	*	0.0018	-0.0493	43.3158
	3′ Templated	0.0684	0.883		0.0174	0.0051	436.5789
	3′ Non-templated (A)	0.0785	0.758		0.0026	-0.0246	67.8947
	3′ Non-templated (U)	0.1034	0.414		0.0003	-0.0445	9.3158
	Substitutions	0.1707	0.028	*	0.0033	-0.0634	82.0526
miR-423-5p	Total	0.0445	0.61		0.0174	0.0088	440.8421
	Canonical	0.0791	0.0519		0.0097	0.0295	249.2105
	5′ Variant	0.0778	0.0588		0.0005	0.0353	12.3684
	3′ Templated	0.0834	0.0345	*	0.006	-0.0302	146.5263
	3′ Non-templated (A)	0.1231	0.000289	***	0.0005	-0.0607	11.7895
	3′ Non-templated (U)	0.1689	1.19E-07	***	0.0002	-0.0982	4.2632
	Substitutions	0.077	0.063		0.0021	0.0219	53.5789
miR-182-5p	Total	0.0806	0.0135	*	0.0132	0.0396	317.8947
	Canonical	0.0698	0.0472	*	0.002	0.0311	49.2632
	5′ Variant	0.1182	4.23E-05	***	0.0018	0.0505	45.1053
	3′ Templated	0.0693	0.0496	*	0.0106	0.0231	252.3158
	3′ Non-templated (A)	0.0626	0.0975		0.0003	-0.026	5.5789
	3′ Non-templated (U)	0.0269	0.944		0.0001	-0.0056	2.9474
	Substitutions	0.0746	0.0277	*	0.0015	0.0323	35.2632

Table Appendix B1. Statistical Values and Read Counts for IsomiR-Target Pairs in K562 cells. Two sets of correlation values were determined, comparing expression of each isomiR category against either the miRNAs canonical targets, or non-targets (control). Distributions of correlation values were compared with the two-sided Kolmogorov–Smirnov test to identify pairs with a significant difference. **Signif:** Significance represented as asterisks - ***:** 0.005 < p.value < 0.05, ****:** 0.0005 < p.value < 0.005, *****:** p.value < 0.0005. **Mean Prop. of Total Reads:** Mean proportion of total reads each category represents across all cells. **Delta Mean:** Difference between the mean of correlation values for isomiR-target gene pairs and mean of correlation values for isomiR-control gene pairs.

miRNA	Category	KS Statistic	P.Value	Signif.	Mean Prop. of Total Reads	Delta Mean	Mean Read Count
miR-122-5p	Total	0.2187	6.96E-10	***	0.4303	0.0255	14103.45
	Canonical	0.3743	0	***	0.0023	0.0791	75.5161
	5′ Variant	0.2633	3.95E-14	***	0.0002	-0.0783	5.7097
	3′ Templated	0.2122	2.51E-09	***	0.2927	0.0252	9597.645
	3′ Non-templated (A)	0.2593	1.03E-13	***	0.1348	0.0349	4418.581
	3′ Non-templated (U)	0.3885	0	***	0.0058	0.0881	191.2258
	Substitutions	0.1578	2.38E-05	***	0.0439	0.0102	1438
let-7i-5p	Total	0.3522	0	***	0.0751	0.1179	3078.655
	Canonical	0.4573	0	***	0.0001	0.1747	2.1034
	5′ Variant	0.354	0	***	0.0576	0.1186	2361.724
	3′ Templated	0.3528	0	***	0.0575	0.118	2358.552
	3′ Non-templated (A)	0.3468	0	***	0.001	0.112	39.0345
	3′ Non-templated (U)	0.3441	0	***	0.0042	0.1109	171.3793
	Substitutions	0.392	0	***	0.0142	0.1419	576.5172
let-7b-5p	Total	0.1726	0	***	0.0578	0.0336	2227.419
	Canonical	0.1372	5.19E-11	***	0.0014	0.0217	52.3871
	5′ Variant	0.1738	0	***	0.0433	0.0341	1678.032
	3′ Templated	0.1472	1.31E-12	***	0.0472	0.0254	1833.258
	3′ Non-templated (A)	0.2097	0	***	0.0024	0.047	89.6774
	3′ Non-templated (U)	0.0629	0.0118	*	0.0198	0.005	798.2903
	Substitutions	0.1332	2.06E-10	***	0.0179	0.0275	712.6129
miR-34a-5p	Total	0.2874	0	***	0.0531	-0.1234	1035.083
	Canonical	0.2827	0	***	0.0012	-0.119	23.0417
	5′ Variant	0.2921	0	***	0.0094	-0.1262	182.7917
	3′ Templated	0.2846	0	***	0.0383	-0.1223	747.4583
	3′ Non-templated (A)	0.2797	0	***	0.0043	-0.1197	84.625
	3′ Non-templated (U)	0.2873	0	***	0.0068	-0.1244	133.9583
	Substitutions	0.2891	0	***	0.0211	-0.1246	411.7083
miR-148a-3p	Total	0.1211	2.24E-07	***	0.0285	-0.0169	1064.065
	Canonical	0.1511	3.04E-11	***	0.01	-0.0256	373.8065
	5′ Variant	0.0948	0.00011	***	0.0034	-0.0276	122.6452
	3′ Templated	0.1431	3.94E-10	***	0.0075	-0.0239	279.7419
	3′ Non-templated (A)	0.1803	7.77E-16	***	0.0009	-0.0417	34.6774
	3′ Non-templated (U)	0.1379	1.97E-09	***	0.0049	-0.0201	182.4839
	Substitutions	0.0956	9.30E-05	***	0.0111	-0.0114	411.7419
miR-206	Total	0.0361	0.476		0.0317	-0.0055	620.4815
	Canonical	0.034	0.555		0.0065	-0.0027	126.5185
	5′ Variant	0.0337	0.565		0.0069	-0.0059	136.4444
	3′ Templated	0.0349	0.52		0.0202	-0.0067	397.1852
	3′ Non-templated (A)	0.0387	0.388		0.0007	-0.0056	14.2222
	3′ Non-templated (U)	0.063	0.0264	*	0.0007	0.0061	13.5185
	Substitutions	0.0348	0.526		0.0077	-0.0104	151.4074

Table Appendix B2. Statistical Values and Read Counts for IsomiR-Target Pairs in Hepatocellular Carcinoma cells. Two sets of correlation values were determined, comparing expression of each isomiR category against either the miRNAs canonical targets, or non-targets (control). Distributions of correlation values were compared with the two-sided Kolmogorov–Smirnov test to identify pairs with a significant difference. **Signif:** Significance represented as asterisks - ***:** 0.005 < p.value < 0.05, ****:** 0.0005 < p.value < 0.005, *****:** p.value < 0.0005. **Mean Prop. of Total Reads:** Mean proportion of total reads each category represents across all cells. **Delta Mean:** Difference between the mean of correlation values for isomiR-target gene pairs and mean of correlation values for isomiR-control gene pairs.

Appendix C – Read Counts for Spliced Exon Junctions in Ago2

Sample	Exon Junction	Associated Transcript	Reads Including Junction	Reads Excluding Junction	PSJ (%)
HeLa	Chr8:140532616-140539319	NM_001164623	12	2359	0.51
	Chr8:140558573-140559419	Unannotated	0	2317	0
	Chr8:140572604-140572811	Unannotated	17	632	2.62
	Chr8:140572933-140635484	XM_017013317	16	784	2
	Chr8:140585312-140588721	Unannotated	2	301	0.66
	Chr8:140585312-140590593	Unannotated	0	303	0
	Chr8:140585312-140596675	XM_011516965	9	294	2.97
	Chr8:140585312-140605796	XM_011516966	32	271	10.56
	Chr8:140585312-140633774	Unannotated	14	291	4.59
	Chr8:140585312-140635484	NM_012154	226	79	74.1
HEK293	Chr8:140532616-140539319	NM_001164623	14	1827	0.76
	Chr8:140558573-140559419	Unannotated	3	2066	0.14
	Chr8:140572604-140572811	Unannotated	4	527	0.75
	Chr8:140572933-140635484	XM_017013317	10	701	1.41
	Chr8:140585312-140588721	Unannotated	4	250	1.57
	Chr8:140585312-140590593	Unannotated	0	254	0
	Chr8:140585312-140596675	XM_011516965	1	253	0.39
	Chr8:140585312-140605796	XM_011516966	23	231	9.06
	Chr8:140585312-140633774	Unannotated	3	251	1.18
	Chr8:140585312-140635484	NM_012154	209	45	82.28
MCF7	Chr8:140532616-140539319	NM_001164623	16	2007	0.79
	Chr8:140558573-140559419	Unannotated	5	1960	0.25
	Chr8:140572604-140572811	Unannotated	6	607	0.98
	Chr8:140572933-140635484	XM_017013317	13	785	1.63
	Chr8:140585312-140588721	Unannotated	0	277	0
	Chr8:140585312-140590593	Unannotated	0	277	0
	Chr8:140585312-140596675	XM_011516965	5	272	1.81
	Chr8:140585312-140605796	XM_011516966	41	236	14.8
	Chr8:140585312-140633774	Unannotated	9	272	3.2
	Chr8:140585312-140635484	NM_012154	206	76	73.05
LNCaP	Chr8:140532616-140539319	NM_001164623	40	4647	0.85
	Chr8:140558573-140559419	Unannotated	12	4467	0.27
	Chr8:140572604-140572811	Unannotated	56	1376	3.91
	Chr8:140572933-140635484	XM_017013317	12	1872	0.64
	Chr8:140585312-140588721	Unannotated	68	764	8.17
	Chr8:140585312-140590593	Unannotated	51	781	6.13
	Chr8:140585312-140596675	XM_011516965	28	804	3.37
	Chr8:140585312-140605796	XM_011516966	71	762	8.52

	Chr8:140585312-140633774	Unannotated	10	842	1.17
	Chr8:140585312-140635484	NM_012154	510	342	59.86
PNT2	Chr8:140532616-140539319	NM_001164623	1	3862	0.03
	Chr8:140558573-140559419	Unannotated	0	4572	0
	Chr8:140572604-140572811	Unannotated	43	1222	3.4
	Chr8:140572933-140635484	XM_017013317	14	2140	0.65
	Chr8:140585312-140588721	Unannotated	17	861	1.94
	Chr8:140585312-140590593	Unannotated	0	878	0
	Chr8:140585312-140596675	XM_011516965	9	869	1.03
	Chr8:140585312-140605796	XM_011516966	31	848	3.53
	Chr8:140585312-140633774	Unannotated	10	878	1.13
	Chr8:140585312-140635484	NM_012154	782	106	88.06
SCR4	Chr8:140532616-140539319	NM_001164623	46	9419	0.49
	Chr8:140558573-140559419	Unannotated	22	8379	0.26
	Chr8:140572604-140572811	Unannotated	41	2593	1.56
	Chr8:140572933-140635484	XM_017013317	63	3500	1.77
	Chr8:140585312-140588721	Unannotated	24	1342	1.76
	Chr8:140585312-140590593	Unannotated	0	1366	0
	Chr8:140585312-140596675	XM_011516965	22	1344	1.61
	Chr8:140585312-140605796	XM_011516966	115	1252	8.41
	Chr8:140585312-140633774	Unannotated	42	1355	3.01
	Chr8:140585312-140635484	NM_012154	1074	326	76.71
SCC38	Chr8:140532616-140539319	NM_001164623	10	3599	0.28
	Chr8:140558573-140559419	Unannotated	11	3843	0.29
	Chr8:140572604-140572811	Unannotated	65	1161	5.3
	Chr8:140572933-140635484	XM_017013317	19	1843	1.02
	Chr8:140585312-140588721	Unannotated	4	835	0.48
	Chr8:140585312-140590593	Unannotated	3	836	0.36
	Chr8:140585312-140596675	XM_011516965	14	825	1.67
	Chr8:140585312-140605796	XM_011516966	58	781	6.91
	Chr8:140585312-140633774	Unannotated	27	825	3.17
	Chr8:140585312-140635484	NM_012154	712	143	83.27
SKN	Chr8:140532616-140539319	NM_001164623	12	5922	0.2
	Chr8:140558573-140559419	Unannotated	3	6014	0.05
	Chr8:140572604-140572811	Unannotated	24	1545	1.53
	Chr8:140572933-140635484	XM_017013317	53	2338	2.22
	Chr8:140585312-140588721	Unannotated	2	984	0.2
	Chr8:140585312-140590593	Unannotated	0	986	0
	Chr8:140585312-140596675	XM_011516965	6	980	0.61
	Chr8:140585312-140605796	XM_011516966	127	859	12.88
	Chr8:140585312-140633774	Unannotated	7	1017	0.68
	Chr8:140585312-140635484	NM_012154	782	242	76.37
THP	Chr8:140532616-140539319	NM_001164623	8	1552	0.51
	Chr8:140558573-140559419	Unannotated	1	1665	0.06

	Chr8:140572604-140572811	Unannotated	7	392	1.75
	Chr8:140572933-140635484	XM_017013317	11	656	1.65
	Chr8:140585312-140588721	Unannotated	0	214	0
	Chr8:140585312-140590593	Unannotated	0	214	0
	Chr8:140585312-140596675	XM_011516965	7	207	3.27
	Chr8:140585312-140605796	XM_011516966	21	193	9.81
	Chr8:140585312-140633774	Unannotated	2	217	0.91
	Chr8:140585312-140635484	NM_012154	171	48	78.08
Foetal Brain	Chr8:140532616-140539319	NM_001164623	0	403	0
	Chr8:140558573-140559419	Unannotated	0	310	0
	Chr8:140572604-140572811	Unannotated	0	76	0
	Chr8:140572933-140635484	XM_017013317	2	105	1.87
	Chr8:140585312-140588721	Unannotated	0	35	0
	Chr8:140585312-140590593	Unannotated	0	35	0
	Chr8:140585312-140596675	XM_011516965	2	33	5.71
	Chr8:140585312-140605796	XM_011516966	0	35	0
	Chr8:140585312-140633774	Unannotated	1	34	2.86
	Chr8:140585312-140635484	NM_012154	28	7	80
Whole Brain	Chr8:140532616-140539319	NM_001164623	1	268	0.37
	Chr8:140558573-140559419	Unannotated	0	172	0
	Chr8:140572604-140572811	Unannotated	0	51	0
	Chr8:140572933-140635484	XM_017013317	0	69	0
	Chr8:140585312-140588721	Unannotated	0	16	0
	Chr8:140585312-140590593	Unannotated	0	16	0
	Chr8:140585312-140596675	XM_011516965	1	15	6.25
	Chr8:140585312-140605796	XM_011516966	0	16	0
	Chr8:140585312-140633774	Unannotated	2	14	12.5
	Chr8:140585312-140635484	NM_012154	11	5	68.75
Bone Marrow	Chr8:140532616-140539319	NM_001164623	0	1587	0
	Chr8:140558573-140559419	Unannotated	53	1199	4.23
	Chr8:140572604-140572811	Unannotated	25	242	9.36
	Chr8:140572933-140635484	XM_017013317	14	457	2.97
	Chr8:140585312-140588721	Unannotated	0	267	0
	Chr8:140585312-140590593	Unannotated	0	267	0
	Chr8:140585312-140596675	XM_011516965	19	248	7.12
	Chr8:140585312-140605796	XM_011516966	53	214	19.85
	Chr8:140585312-140633774	Unannotated	0	290	0
	Chr8:140585312-140635484	NM_012154	181	109	62.41
Heart	Chr8:140532616-140539319	NM_001164623	22	2590	0.84
	Chr8:140558573-140559419	Unannotated	0	2178	0
	Chr8:140572604-140572811	Unannotated	16	485	3.19
	Chr8:140572933-140635484	XM_017013317	1	840	0.12
	Chr8:140585312-140588721	Unannotated	2	249	0.8
	Chr8:140585312-140590593	Unannotated	0	251	0

	Chr8:140585312-140596675	XM_011516965	3	248	1.2
	Chr8:140585312-140605796	XM_011516966	6	245	2.39
	Chr8:140585312-140633774	Unannotated	7	244	2.79
	Chr8:140585312-140635484	NM_012154	232	19	92.43
Kidney	Chr8:140532616-140539319	NM_001164623	4	734	0.54
	Chr8:140558573-140559419	Unannotated	1	695	0.14
	Chr8:140572604-140572811	Unannotated	2	196	1.01
	Chr8:140572933-140635484	XM_017013317	15	245	5.77
	Chr8:140585312-140588721	Unannotated	2	92	2.13
	Chr8:140585312-140590593	Unannotated	0	94	0
	Chr8:140585312-140596675	XM_011516965	0	94	0
	Chr8:140585312-140605796	XM_011516966	5	89	5.32
	Chr8:140585312-140633774	Unannotated	0	94	0
	Chr8:140585312-140635484	NM_012154	72	22	76.6
Skeletal Muscle	Chr8:140532616-140539319	NM_001164623	8	2222	0.36
	Chr8:140558573-140559419	Unannotated	7	2099	0.33
	Chr8:140572604-140572811	Unannotated	48	544	8.11
	Chr8:140572933-140635484	XM_017013317	5	873	0.57
	Chr8:140585312-140588721	Unannotated	0	258	0
	Chr8:140585312-140590593	Unannotated	0	258	0
	Chr8:140585312-140596675	XM_011516965	0	258	0
	Chr8:140585312-140605796	XM_011516966	0	258	0
	Chr8:140585312-140633774	Unannotated	2	256	0.78
	Chr8:140585312-140635484	NM_012154	251	7	97.29
Testes	Chr8:140532616-140539319	NM_001164623	0	369	0
	Chr8:140558573-140559419	Unannotated	0	239	0
	Chr8:140572604-140572811	Unannotated	0	10	0
	Chr8:140572933-140635484	XM_017013317	0	48	0
	Chr8:140585312-140588721	Unannotated	0	12	0
	Chr8:140585312-140590593	Unannotated	0	12	0
	Chr8:140585312-140596675	XM_011516965	0	12	0
	Chr8:140585312-140605796	XM_011516966	1	11	8.33
	Chr8:140585312-140633774	Unannotated	1	11	8.33
	Chr8:140585312-140635484	NM_012154	10	2	83.33

Table Appendix C. Read Counts and Percent Spliced with Junction (PSJ) values for Spliced Exon Junctions in Ago2. Includes 16 tissues and cell lines, with all junctions unique to annotated transcripts and the top 5 unannotated junctions by mean read count across samples.

Appendix D – Western Blots for the Immunoprecipitation of Ago2 Isoforms

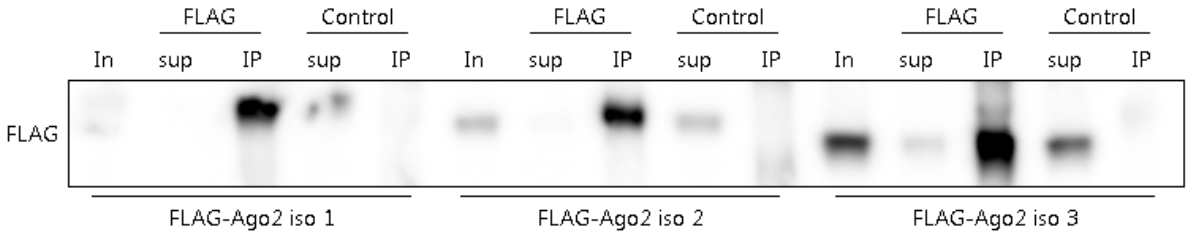

Figure Appendix D. Western blots confirm enrichment of FLAG antibodies in FLAG immunoprecipitate compared to control for all three Ago2 isoforms. Equal concentrations of protein were used across FLAG and control supernatant and IP samples. Mouse IgG antibody was used as a control. RNA isolated from immunoprecipitates were subsequently used for small RNA sequencing. **In:** Input (cell lysate). **sup:** supernatant of immunoprecipitated lysate. **IP:** Immunoprecipitated sample targeted for enrichment of antibody epitope. **Iso:** isoform.

Bibliography

1. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014).

2. Nguyen, T. A. *et al.* Functional anatomy of the human microprocessor. *Cell* **161**, 1374–1387 (2015).

3. Guo, L., Liang, T., Yu, J. & Zou, Q. A comprehensive analysis of miRNA/isomiR expression with gender difference. *PLoS One* **11**, 1–12 (2016).

4. Han, J. *et al.* The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016–3027 (2004).

5. Han, J. *et al.* Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell* **125**, 887–901 (2006).

6. Kwon, S. C. *et al.* Molecular Basis for the Single-Nucleotide Precision of Primary microRNA Processing. *Mol. Cell* **73**, 505-518.e5 (2018).

7. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).

8. Kim, Y. K., Kim, B. & Kim, V. N. Re-evaluation of the roles of DROSHA, Exportin 5, and DICER in microRNA biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E1881–E1889 (2016).

9. Wilson, R. C. *et al.* Dicer-TRBP complex formation ensures accurate mammalian MicroRNA biogenesis. *Mol. Cell* **57**, 397–407 (2015).

10. Lee, Y. *et al.* The role of PACT in the RNA silencing pathway. *EMBO J.* **25**, 522–532 (2006).

11. Diederichs, S. & Haber, D. a. Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell* **131**, 1097–108 (2007).

12. Kawamata, T. & Tomari, Y. Making RISC. *Trends Biochem. Sci.* **35**, 368–376 (2010).

13. Kobayashi, H. & Tomari, Y. RISC assembly: Coordination between small RNAs and Argonaute proteins. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859**, 71–81 (2016).

14. Iwasaki, S. *et al.* Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol. Cell* **39**, 292–9 (2010).

15. Kwak, P. B. & Tomari, Y. The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat. Struct. Mol. Biol.* **19**, 145–151 (2012).

16. Westholm, J. O. & Lai, E. C. Mirtrons: MicroRNA biogenesis via splicing. *Biochimie* **93**, 1897–1904 (2011).

17. Maute, R. L. *et al.* tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. *Proc. Natl. Acad. Sci.* **110**, 1404–1409 (2013).

18. Ender, C. *et al.* A Human snoRNA with MicroRNA-Like Functions. *Mol. Cell* **32**, 519–528 (2008).

19. Lemus-Diaz, N., Ferreira, R. R., Bohnsack, K. E., Gruber, J. & Bohnsack, M. T. The human box C/D snoRNA U3 is a miRNA source and miR-U3 regulates expression of sortin nexin 27. *Nucleic Acids Res.* **48**, 8074–8089 (2020).

20. Cheloufi, S., Dos Santos, C. O., Chong, M. M. W. & Hannon, G. J. A dicer-independent miRNA

biogenesis pathway that requires Ago catalysis. *Nature* **465**, 584–589 (2010).

21. Kretov, D. A. *et al.* Ago2-Dependent Processing Allows miR-451 to Evade the Global MicroRNA Turnover Elicited during Erythropoiesis. *Mol. Cell* **78**, 317-328.e6 (2020).

22. Kornblihtt, A. R. *et al.* Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14**, 153–165 (2013).

23. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).

24. Dai, L. *et al.* Cytoplasmic Drosha activity generated by alternative splicing. *Nucleic Acids Res.* **44**, 10454–10466 (2016).

25. Dai, L. *et al.* Novel, abundant Drosha isoforms are deficient in miRNA processing in cancer cells. *RNA Biol.* **17**, 1603–1612 (2020).

26. Potenza, N. *et al.* A novel splice variant of the human dicer gene is expressed in neuroblastoma cells. *FEBS Lett.* **584**, 3452–3457 (2010).

27. Flemr, M. *et al.* A Retrotransposon-Driven Dicer Isoform Directs Endogenous Small Interfering RNA Production in Mouse Oocytes. *Cell* **155**, 807–816 (2013).

28. Link, S., Grund, S. E. & Diederichs, S. Alternative splicing affects the subcellular localization of Drosha. *Nucleic Acids Res.* **44**, 5330–5343 (2016).

29. Martinez, I. *et al.* An Exportin-1-dependent microRNA biogenesis pathway during human cell quiescence. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4961–E4970 (2017).

30. Grelier, G. *et al.* Prognostic value of Dicer expression in human breast cancers and association with the mesenchymal phenotype. *Br. J. Cancer* **101**, 673–683 (2009).

31. Cantini, L. P. *et al.* Identification and characterization of Dicer1e, a Dicer1 protein variant, in oral cancer cells. *Mol. Cancer* **13**, 1–15 (2014).

32. Irvin-Wilson, C. V. & Chaudhuri, G. Alternative initiation and splicing in dicer gene expression in human breast cells. *Breast Cancer Res.* **7**, 563–569 (2005).

33. Hinkal, G. W., Grelier, G., Puisieux, A. & Moyret-Lalle, C. Complexity in the regulation of dicer expression: Dicer variant proteins are differentially expressed in epithelial and mesenchymal breast cancer cells and decreased during EMT. *Br. J. Cancer* **104**, 387–388 (2011).

34. Singh, A. *et al.* Let-7a-regulated translational readthrough of mammalian AGO 1 generates a micro RNA pathway inhibitor . *EMBO J.* **38**, 1–20 (2019).

35. Ghosh, S. *et al.* Prevention of dsRNA-induced interferon signaling by AGO1x is linked to breast cancer cell proliferation. *EMBO J.* **39**, 1–23 (2020).

36. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).

37. Benetatos, L. *et al.* The microRNAs within the DLK1-DIO3 genomic region: Involvement in disease pathogenesis. *Cell. Mol. Life Sci.* **70**, 795–814 (2013).

38. Bortolin-Cavaille, M. L., Dance, M., Weber, M. & Cavaille, J. C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts. *Nucleic Acids Res.* **37**, 3464–3473 (2009).

39. Kabekkodu, S. P. *et al.* Clustered miRNAs and their role in biological functions and diseases.

Biol. Rev. **93**, 1955–1986 (2018).

40. Altuvia, Y. Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res.* **33**, 2697–2706 (2005).

41. Mohammed, J., Siepel, A. & Lai, E. C. Diverse modes of evolutionary emergence and flux of conserved microRNA clusters. *RNA* **20**, 1850–1863 (2014).

42. Sylvestre, Y. *et al.* An E2F/miR-20a autoregulatory feedback loop. *J. Biol. Chem.* **282**, 2135–2143 (2007).

43. Luo, Z. *et al.* Regulation of the imprinted Dlk1-Dio3 locus by allele-specific enhancer activity. *Genes Dev.* **30**, 92–101 (2016).

44. Cavaille, J. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum. Mol. Genet.* **11**, 1527–1538 (2002).

45. Jiang, W. & Chen, L. Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. *Comput. Struct. Biotechnol. J.* **19**, 183–195 (2021).

46. Morin, R. D. *et al.* Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* **18**, 610–621 (2008).

47. Newman, M. A., Mani, V. & Hammond, S. M. Deep sequencing of microRNA precursors reveals extensive 3′ end modification. *Rna* **17**, 1795–1803 (2011).

48. Menezes, M. R., Balzeau, J. & Hagan, J. P. 3′ RNA Uridylation in Epitranscriptomics, Gene Regulation, and Disease. *Front. Mol. Biosci.* **5**, 1–20 (2018).

49. Kim, Y.-K., Kim, B. & Kim, V. N. Re-evaluation of the roles of DROSHA , Exportin 5 , and DICER in microRNA biogenesis. *Proc. Natl. Acad. Sci.* **113**, E1881–E1889 (2016).

50. Bofill-De Ros, X. *et al.* Structural Differences between Pri-miRNA Paralogs Promote Alternative Drosha Cleavage and Expand Target Repertoires. *Cell Rep.* **26**, 447-459.e4 (2019).

51. Heo, I. *et al.* Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell* **151**, 521–532 (2012).

52. Kim, B. *et al.* TUT7 controls the fate of precursor microRNAs by using three different uridylation mechanisms. *EMBO J.* **34**, 1801–1815 (2015).

53. Heo, I. *et al.* TUT4 in Concert with Lin28 Suppresses MicroRNA Biogenesis through Pre-MicroRNA Uridylation. *Cell* **138**, 696–708 (2009).

54. Jones, M. R. *et al.* Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat. Cell Biol.* **11**, 1157–1163 (2009).

55. Tan, G. C. *et al.* 5′ isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.* **42**, 9424–9435 (2014).

56. D'Ambrogio, A., Gu, W., Udagawa, T., Mello, C. C. & Richter, J. D. Specific miRNA Stabilization by Gld2-Catalyzed Monoadenylation. *Cell Rep.* **2**, 1537–1545 (2012).

57. Thornton, J. E. *et al.* Selective microRNA uridylation by Zcchc6 (TUT7) and Zcchc11 (TUT4). *Nucleic Acids Res.* **42**, 11777–11791 (2014).

58. Baccarini, A. *et al.* Kinetic analysis reveals the fate of a MicroRNA following target regulation in mammalian cells. *Curr. Biol.* **21**, 369–376 (2011).

59. Ameres, S. L. *et al.* Target RNA-directed trimming and tailing of small silencing RNAs. *Science (80-.).* **328**, 1534–1539 (2010).

60. Stadler, M. B. *et al.* Potent degradation of neuronal miRNAs induced by highly complementary targets. *EMBO Rep.* **16**, 500–511 (2015).

61. Haas, G. *et al.* Identification of factors involved in target RNA-directed microRNA degradation. *Nucleic Acids Res.* **44**, 2873–2887 (2016).

62. Simeone, I. *et al.* Endogenous transcripts control miRNA levels and activity in mammalian cells by target-directed miRNA degradation. *Nat. Commun.* **9**, (2018).

63. Hutvágner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–60 (2002).

64. Gu, S. & Kay, M. A. How do miRNAs mediate translational repression? *Silence* **1**, 1–5 (2010).

65. Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W. & Pasquinelli, A. E. Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Mol. Cell* **64**, 320–333 (2016).

66. Schirle, N. T., Sheu-Gruttadauria, J. & MacRae, I. J. Structural basis for microRNA targeting. *Science (80-.).* **346**, 608–613 (2014).

67. Hutvagner, G. & Simard, M. J. Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.* **9**, 22–32 (2008).

68. Müller, M., Fazi, F. & Ciaudo, C. Argonaute Proteins: From Structure to Function in Development and Pathological Cell Fate Determination. *Front. Cell Dev. Biol.* **7**, 1–10 (2020).

69. Matsui, M., Li, L., Janowski, B. A. & Corey, D. R. Reduced Expression of Argonaute 1, Argonaute 2, and TRBP Changes Levels and Intracellular Distribution of RNAi Factors. *Sci. Rep.* **5**, 1–11 (2015).

70. Lykke-Andersen, K. *et al.* Maternal Argonaute 2 Is Essential for Early Mouse Development at the Maternal-Zygotic Transition. *Mol. Biol. Cell* **19**, 4383–4392 (2008).

71. Park, M. S., Sim, G., Kehling, A. C. & Nakanishi, K. Human Argonaute2 and Argonaute3 are catalytically activated by different lengths of guide RNA. *Proc. Natl. Acad. Sci.* **117**, 28576–28578 (2020).

72. Parker, J. S., Roe, S. M. & Barford, D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**, 663–666 (2005).

73. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).

74. Iwakawa, H. oki & Tomari, Y. The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends Cell Biol.* **25**, 651–665 (2015).

75. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).

76. Braun, J. E., Huntzinger, E. & Izaurralde, E. A molecular link between miRISCs and deadenylases provides new insight into the mechanism of gene silencing by microRNAs. *Cold Spring Harb. Perspect. Biol.* **4**, (2012).

77. Takimoto, K., Wakiyama, M. & Yokoyama, S. Mammalian GW182 contains multiple Argonaute-binding sites and functions in microRNA-mediated translational repression. *RNA* **15**, 1078–1089 (2009).

78. Mathonnet, G. *et al.* MicroRNA Inhibition of Translation Initiation in Vitro by Targeting the Cap-Binding Complex eIF4F. *Science (80-.).* **317**, 1764–1767 (2007).

79. Fukaya, T., Iwakawa, H. & Tomari, Y. MicroRNAs Block Assembly of eIF4F Translation Initiation Complex in Drosophila. *Mol. Cell* **56**, 67–78 (2014).

80. Wahle, E. & Winkler, G. S. RNA decay machines: Deadenylation by the Ccr4–Not and Pan2–Pan3 complexes. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829**, 561–570 (2013).

81. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).

82. Calin, G. A. *et al.* Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 15524–9 (2002).

83. Calin, G. A. *et al.* Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl. Acad. Sci.* **101**, 2999–3004 (2004).

84. Blenkiron, C. *et al.* MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol.* **8**, R214 (2007).

85. LI, X., YANG, H., TIAN, Q., LIU, Y. & WENG, Y. Upregulation of microRNA-17-92 cluster associates with tumor progression and prognosis in osteosarcoma. *Neoplasma* **61**, 453–460 (2014).

86. Leichter, A. L., Sullivan, M. J., Eccles, M. R. & Chatterjee, A. MicroRNA expression patterns and signalling pathways in the development and progression of childhood solid tumours. *Mol. Cancer* **16**, 1–17 (2017).

87. De Preter, K. *et al.* miRNA expression profiling enables risk stratification in archived and fresh neuroblastoma tumor samples. *Clin. Cancer Res.* **17**, 7684–7692 (2011).

88. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).

89. Malumbres, M. MiRNAs and cancer: An epigenetics view. *Mol. Aspects Med.* **34**, 863–874 (2013).

90. Esquela-Kerscher, A. & Slack, F. J. Oncomirs - MicroRNAs with a role in cancer. *Nat. Rev. Cancer* **6**, 259–269 (2006).

91. Williams, M., Cheng, Y. Y., Blenkiron, C. & Reid, G. Exploring Mechanisms of MicroRNA Downregulation in Cancer. *MicroRNA* **6**, 2–16 (2017).

92. Di Leva, G., Garofalo, M. & Croce, C. M. MicroRNAs in Cancer. *Annu. Rev. Pathol. Mech. Dis.* **9**, 287–314 (2014).

93. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).

94. Asadzadeh, Z. *et al.* microRNAs in cancer stem cells: Biology, pathways, and therapeutic opportunities. *J. Cell. Physiol.* **234**, 10002–10017 (2019).

95. Hata, A. & Kashima, R. Dysregulation of microRNA biogenesis machinery in cancer. *Crit. Rev. Biochem. Mol. Biol.* **51**, 121–134 (2016).

96. Wen, J., Lv, Z., Ding, H., Fang, X. & Sun, M. Association of miRNA biosynthesis genes DROSHA and DGCR8 polymorphisms with cancer susceptibility: A systematic review and meta-analysis. *Biosci. Rep.* **38**, (2018).

97. Czubak, K. *et al.* High copy number variation of cancer-related microRNA genes and frequent

amplification of DICER1 and DROSHA in lung cancer. *Oncotarget* **6**, 23399–23416 (2015).

98. Guo, Y. *et al.* Silencing the double-stranded RNA binding protein DGCR8 inhibits ovarian cancer cell proliferation, migration, and invasion. *Pharm. Res.* **32**, 769–778 (2015).

99. Belair, C. D. *et al.* DGCR8 is essential for tumor progression following PTEN loss in the prostate. *EMBO Rep.* **16**, 1219–32 (2015).

100. Davis, B. N., Hilyard, A. C., Lagna, G. & Hata, A. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **454**, 56–61 (2008).

101. Wu, K., He, J., Pu, W. & Peng, Y. The Role of Exportin-5 in MicroRNA Biogenesis and Cancer. *Genomics. Proteomics Bioinformatics* **16**, 120–126 (2018).

102. Li, Y., Wang, X., He, B., Cai, H. & Gao, Y. Downregulation and tumor-suppressive role of XPO5 in hepatocellular carcinoma. *Mol. Cell. Biochem.* **415**, 197–205 (2016).

103. Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat. Genet.* **39**, 673–677 (2007).

104. Lambertz, I. *et al.* Monoallelic but not biallelic loss of Dicer1 promotes tumorigenesis in vivo. *Cell Death Differ.* **17**, 633–641 (2010).

105. Robertson, J. C., Jorcyk, C. L. & Oxford, J. T. DICER1 syndrome: DICER1 mutations in rare cancers. *Cancers (Basel).* **10**, 1–17 (2018).

106. Solarski, M. *et al.* DICER1 gene mutations in endocrine tumors. *Endocr. Relat. Cancer* **25**, R197–R208 (2018).

107. Ye, Z. L., Jin, H. J. & Qian, Q. J. Argonaute 2: A novel rising star in cancer research. *J. Cancer* **6**, 877–882 (2015).

108. Cheng, N., Li, Y. & Han, Z.-G. Argonaute2 promotes tumor metastasis by way of up-regulating focal adhesion kinase expression in hepatocellular carcinoma. *Hepatology* **57**, 1906–1918 (2013).

109. Völler, D., Reinders, J., Meister, G. & Bosserhoff, A.-K. Strong reduction of AGO2 expression in melanoma and cellular consequences. *Br. J. Cancer* **109**, 3116–3124 (2013).

110. Shen, J. *et al.* EGFR modulates microRNA maturation in response to hypoxia through phosphorylation of AGO2. *Nature* **497**, 383–387 (2013).

111. Zhang, H. *et al.* Acetylation of AGO2 promotes cancer progression by increasing oncogenic miR-19b biogenesis. *Oncogene* **38**, 1410–1431 (2019).

112. Torrezan, G. T. *et al.* Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour. *Nat. Commun.* **5**, 4039 (2014).

113. Wu, M. K. *et al.* Biallelic DICER1 mutations occur in Wilms tumours. *J. Pathol.* **230**, 154–164 (2013).

114. Rakheja, D. *et al.* Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. *Nat. Commun.* **2**, 1–10 (2014).

115. Walz, A. L. *et al.* Recurrent DGCR8, DROSHA, and SIX Homeodomain Mutations in Favorable Histology Wilms Tumors. *Cancer Cell* **27**, 286–297 (2015).

116. Wegert, J. *et al.* Mutations in the SIX1/2 Pathway and the DROSHA/DGCR8 miRNA Microprocessor Complex Underlie High-Risk Blastemal Type Wilms Tumors. *Cancer Cell* **27**,

298–311 (2015).

117. Gadd, S. *et al.* A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nat. Genet.* **49**, 1487–1494 (2017).

118. Diederichs, S. & Haber, D. A. Sequence variations of microRNAs in human cancer: Alterations in predicted secondary structure do not affect processing. *Cancer Res.* **66**, 6097–6104 (2006).

119. Maqbool, R., Ismail, R. & Hussain, M.-. Mutations in MicroRNA Genes and Their Binding Sites are Infrequently Associated with Human Colorectal Cancer in the Kashmiri Population. *MicroRNA* **2**, 219–224 (2014).

120. Moszyńska, A., Gebert, M., Collawn, J. F. & Bartoszewski, R. SNPs in microRNA target sites and their potential role in human disease. *Open Biol.* **7**, 170019 (2017).

121. Lopes-Ramos, C. M. *et al.* E2F1 somatic mutation within miRNA target site impairs gene regulation in colorectal cancer. *PLoS One* **12**, e0181153 (2017).

122. Agueli, C. *et al.* 14q32/miRNA clusters loss of heterozygosity in acute lymphoblastic leukemia is associated with up-regulation of BCL11a. *Am. J. Hematol.* **85**, 575–578 (2010).

123. Anauate, A. C. *et al.* Analysis of 8q24.21 miRNA cluster expression and copy number variation in gastric cancer. *Future Med. Chem.* **11**, 947–958 (2019).

124. Xia, E. *et al.* MicroRNA induction by copy number gain is associated with poor outcome in squamous cell carcinoma of the lung. *Sci. Rep.* **8**, 15363 (2018).

125. Weber, B., Stresemann, C., Brueckner, B. & Lyko, F. Methylation of human MicroRNA genes in normal and neoplastic cells. *Cell Cycle* **6**, 1001–1005 (2007).

126. Saito, Y. *et al.* Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell* **9**, 435–443 (2006).

127. Lujambio, A. *et al.* Genetic Unmasking of an Epigenetically Silenced microRNA in Human Cancer Cells. *Cancer Res.* **67**, 1424–1429 (2007).

128. Lujambio, A. *et al.* A microRNA DNA methylation signature for human cancer metastasis. *Proc. Natl. Acad. Sci.* **105**, 13556–13561 (2008).

129. Ortiz, I. M. D. P. *et al.* Loss of DNA methylation is related to increased expression of miR-21 and miR-146b in papillary thyroid carcinoma. *Clin. Epigenetics* **10**, 144 (2018).

130. Cao, Q. *et al.* Coordinated Regulation of Polycomb Group Complexes through microRNAs in Cancer. *Cancer Cell* **20**, 187–199 (2011).

131. Au, S. L. K. *et al.* Enhancer of zeste homolog 2 epigenetically silences multiple tumor suppressor microRNAs to promote liver cancer metastasis. *Hepatology* **56**, 622–631 (2012).

132. Yamagishi, M. *et al.* Polycomb-Mediated Loss of miR-31 Activates NIK-Dependent NF-κB Pathway in Adult T Cell Leukemia and Other Cancers. *Cancer Cell* **21**, 121–135 (2012).

133. Sampath, D. *et al.* Histone deacetylases mediate the silencing of miR-15a, miR-16, and miR-29b in chronic lymphocytic leukemia. *Blood* **119**, 1162–1172 (2012).

134. Lee, T. I. & Young, R. A. Transcriptional Regulation and Its Misregulation in Disease. *Cell* **152**, 1237–1251 (2013).

135. O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V. & Mendell, J. T. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435**, 839–843 (2005).

136. Wang, B. *et al.* Reciprocal regulation of microRNA-122 and c-Myc in hepatocellular cancer: Role of E2F1 and transcription factor dimerization partner 2. *Hepatology* **59**, 555–566 (2014).

137. Hermeking, H. p53 Enters the MicroRNA World. *Cancer Cell* **12**, 414–418 (2007).

138. Suzuki, H. I. *et al.* Modulation of microRNA processing by p53. *Nature* **460**, 529–533 (2009).

139. Hermeking, H. MicroRNAs in the p53 network: Micromanagement of tumour suppression. *Nat. Rev. Cancer* **12**, 613–626 (2012).

140. Smith, C. M., Catchpoole, D. & Hutvagner, G. Non-Coding RNAs in Pediatric Solid Tumors. *Front. Genet.* **10**, 1–18 (2019).

141. Wu, X. *et al.* Comprehensive expression analysis of miRNA in breast cancer at the miRNA and isomiR levels. *Gene* **557**, 195–200 (2015).

142. Telonis, A. G. *et al.* Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. *Nucleic Acids Res.* **45**, 2973–2985 (2016).

143. Salem, O. *et al.* The highly expressed 5'isomiR of hsa-miR-140-3p contributes to the tumor-suppressive effects of miR-140 by reducing breast cancer proliferation and migration. *BMC Genomics* **17**, 1–16 (2016).

144. Wang, J., Chen, J. & Sen, S. MicroRNA as Biomarkers and Diagnostics. *J. Cell. Physiol.* **231**, 25–30 (2016).

145. Hamam, R. *et al.* Circulating microRNAs in breast cancer: novel diagnostic and prognostic biomarkers. *Cell Death Dis.* **8**, e3045 (2017).

146. Wang, H., Peng, R., Wang, J., Qin, Z. & Xue, L. Circulating microRNAs as potential cancer biomarkers: The advantage and disadvantage. *Clin. Epigenetics* **10**, 1–10 (2018).

147. Tellez-Gabriel, M., Ory, B., Lamoureux, F., Heymann, M. F. & Heymann, D. Tumour heterogeneity: The key advantages of single-cell analysis. *Int. J. Mol. Sci.* **17**, (2016).

148. Wu, F. *et al.* Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* **12**, 1–11 (2021).

149. Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science (80-.).* **355**, (2017).

150. Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160 (2014).

151. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* **29**, 1120–1127 (2011).

152. Liu, J., Dang, H. & Wang, X. W. The significance of intertumor and intratumor heterogeneity in liver cancer. *Exp. Mol. Med.* **50**, e416-8 (2018).

153. Chen, R., Smith-Cohn, M., Cohen, A. L. & Colman, H. Glioma Subclassifications and Their Clinical Significance. *Neurotherapeutics* vol. 14 284–297 (2017).

154. Schmidt, F. & Efferth, T. Tumor heterogeneity, single-cell sequencing, and drug resistance. *Pharmaceuticals* **9**, (2016).

155. Lawson, D. A., Kessenbrock, K., Davis, R. T., Pervolarakis, N. & Werb, Z. Tumour heterogeneity and metastasis at single-cell resolution. *Nat. Cell Biol.* **20**, 1349–1360 (2018).

156. Saunders, N. A. *et al.* Role of intratumoural heterogeneity in cancer drug resistance: Molecular and clinical perspectives. *EMBO Mol. Med.* **4**, 675–684 (2012).

157. Marusyk, A., Janiszewska, M. & Polyak, K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell* **37**, 471–484 (2020).

158. Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: Implications for targeted therapeutics. *Br. J. Cancer* **108**, 479–485 (2013).

159. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* **15**, 81–94 (2018).

160. Li, T. *et al.* MiR-452 promotes an aggressive colorectal cancer phenotype by regulating a Wnt/β-catenin positive feedback loop. *J. Exp. Clin. Cancer Res.* **37**, 1–15 (2018).

161. Piccart-Gebhart, M. J. *et al.* Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer. *N. Engl. J. Med.* **353**, 1659–1672 (2005).

162. Fidler, I. J. & Hart, I. R. Biological Diversity in Metastatic Neoplasms: Origins and Implications. *Science (80-.).* **217**, 998–1003 (1982).

163. Caiado, F., Silva-Santos, B. & Norell, H. Intra-tumour heterogeneity – going beyond genetics. *FEBS J.* **283**, 2245–2258 (2016).

164. Rich, J. N. Cancer stem cells. *Medicine (Baltimore).* **95**, S2–S7 (2016).

165. Thankamony, A. P., Saxena, K., Murali, R., Jolly, M. K. & Nair, R. Cancer Stem Cell Plasticity – A Deadly Deal. *Front. Mol. Biosci.* **7**, (2020).

166. González-Silva, L., Quevedo, L. & Varela, I. Tumor Functional Heterogeneity Unraveled by scRNA-seq Technologies. *Trends in Cancer* **6**, 13–19 (2020).

167. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).

168. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, (2018).

169. Hayden, E. C. The $1,000 genome 2006. *Nature* **507**, 294–295 (2014).

170. Preston, J., VanZeeland, A. & Peiffer, D. A. Innovation at Illumina: The road to the $600 human genome. *Nat. Portf.* (2021).

171. Lee, J., Hyeon, D. Y. & Hwang, D. Single-cell multiomics: technologies and data analysis methods. *Exp. Mol. Med.* **52**, 1428–1442 (2020).

172. Zhang, Y. *et al.* Single-cell RNA sequencing in cancer research. *J. Exp. Clin. Cancer Res.* **40**, 1–17 (2021).

173. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–95 (2011).

174. Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879-893.e13 (2018).

175. Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422–426 (2015).

176. Morita, K. *et al.* Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat. Commun.* **11**, 5327 (2020).

177. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).

178. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* **9**, 1–12 (2017).

179. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).

180. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Rep.* **2**, 666–673 (2012).

181. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).

182. Zilionis, R. *et al.* Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017).

183. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).

184. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).

185. Cusanovich, D. A. *et al.* Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science (80-.).* **348**, 910–914 (2015).

186. Guo, H. *et al.* Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* **23**, 2126–2135 (2013).

187. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).

188. Faridani, O. R. *et al.* Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.* **34**, 1264–1266 (2016).

189. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).

190. Macaulay, I. C. *et al.* G&T-seq: Parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12**, 519–522 (2015).

191. Wang, N. *et al.* Single-cell microRNA-mRNA co-sequencing reveals non-genetic heterogeneity and mechanisms of microRNA regulation. *Nat. Commun.* **10**, 1–12 (2019).

192. Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* **13**, 229–232 (2016).

193. Khnouf, R. *et al.* Efficient Production of On-Target Reads for Small RNA Sequencing of Single Cells Using Modified Adapters. *Anal. Chem.* **90**, 12609–12615 (2018).

194. Xiao, Z. *et al.* Holo-Seq: Single-cell sequencing of holo-transcriptome. *Genome Biol.* **19**, 1–22 (2018).

195. Hücker, S. M. *et al.* Single-cell microRNA sequencing method comparison and application to cell lines and circulating lung tumor cells. *Nat. Commun.* **12**, 1–13 (2021).

196. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of

Actions, and Circulation. *Front. Endocrinol. (Lausanne).* **9**, (2018).

197. Shim, J. & Nam, J. W. The expression and functional roles of microRNAs in stem cell differentiation. *BMB Rep.* **49**, 3–10 (2016).

198. Bayraktar, R., Van Roosbroeck, K. & Calin, G. A. Cell-to-cell communication: microRNAs as hormones. *Mol. Oncol.* **11**, 1673–1686 (2017).

199. Liang, T., Yu, J., Liu, C. & Guo, L. IsomiR expression patterns in canonical and Dicer-independent microRNAs. *Mol. Med. Rep.* **15**, 1071–1078 (2017).

200. Lan, C., Peng, H., McGowan, E. M., Hutvagner, G. & Li, J. An isomiR expression panel based novel breast cancer classification approach using improved mutual information. *BMC Med. Genomics* **11**, (2018).

201. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

202. Smith, C. M. & Hutvagner, G. A comparative analysis of single cell small RNA sequencing data reveals heterogeneous isomiR expression and regulation. *Sci. Rep.* **12**, 2834 (2022).

203. Friedmann-Morvinski, D. Glioblastoma heterogeneity and cancer cell plasticity. *Crit. Rev. Oncog.* **19**, 327–336 (2014).

204. Tirosh, I. & Suvà, M. L. Tackling the Many Facets of Glioblastoma Heterogeneity. *Cell Stem Cell* **26**, 303–304 (2020).

205. Parker, N. R., Khong, P., Parkinson, J. F., Howell, V. M. & Wheeler, H. R. Molecular heterogeneity in glioblastoma: Potential clinical implications. *Front. Oncol.* **5**, 1–9 (2015).

206. Verhaak, R. G. W. *et al.* Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).

207. Wang, Q. *et al.* Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* **32**, 42-56.e6 (2017).

208. Zhang, P., Xia, Q., Liu, L., Li, S. & Dong, L. Current Opinion on Molecular Characterization for GBM Classification in Guiding Clinical Diagnosis, Prognosis, and Therapy. *Front. Mol. Biosci.* **7**, 1–13 (2020).

209. Singh, S. K. *et al.* Identification of human brain tumour initiating cells. *Nature* **432**, 396–401 (2004).

210. Anido, J. *et al.* TGF-β Receptor Inhibitors Target the CD44high/Id1high Glioma-Initiating Cell Population in Human Glioblastoma. *Cancer Cell* **18**, 655–668 (2010).

211. Hattermann, K. *et al.* Stem cell markers in glioma progression and recurrence. *Int. J. Oncol.* **49**, 1899–1910 (2016).

212. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (80-.).* **344**, 1396–1401 (2014).

213. Darmanis, S. *et al.* Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Rep.* **21**, 1399–1410 (2017).

214. Bhaduri, A. *et al.* Outer Radial Glia-like Cancer Stem Cells Contribute to Heterogeneity of Glioblastoma. *Cell Stem Cell* **26**, 48-63.e6 (2020).

215. Brett, J. O., Renault, V. M., Rafalski, V. A., Webb, A. E. & Brunet, A. The microRNA cluster miR-106b~25 regulates adult neural stem/progenitor cell proliferation and neuronal differentiation. *Aging (Albany. NY).* **3**, 108–124 (2011).

216. Barca-Mayo, O. & Lu, Q. R. Fine-Tuning Oligodendrocyte Development by microRNAs. *Front. Neurosci.* **6**, (2012).

217. Smirnova, L. *et al.* Regulation of miRNA expression during neural cell specification. *Eur. J. Neurosci.* **21**, 1469–1477 (2005).

218. Lavon, I. *et al.* Gliomas display a microRNA expression profile reminiscent of neural precursor cells. *Neuro. Oncol.* **12**, 422–433 (2010).

219. Ciafrè, S. A. *et al.* Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochem. Biophys. Res. Commun.* **334**, 1351–1358 (2005).

220. Chan, J. A., Krichevsky, A. M. & Kosik, K. S. MicroRNA-21 Is an Antiapoptotic Factor in Human Glioblastoma Cells. *Cancer Res.* **65**, 6029–6033 (2005).

221. Rao, S. A., Santosh, V. & Somasundaram, K. Genome-wide expression profiling identifies deregulated miRNAs in malignant astrocytoma. *Mod. Pathol.* **23**, 1404–1417 (2010).

222. Yeh, M. *et al.* MicroRNA-138 suppresses glioblastoma proliferation through downregulation of CD44. *Sci. Rep.* **11**, 1–11 (2021).

223. Xu, B. *et al.* MicroRNAs involved in the EGFR pathway in glioblastoma. *Biomed. Pharmacother.* **134**, 111115 (2021).

224. Chen, M., Medarova, Z. & Moore, A. Role of microRNAs in glioblastoma. *Oncotarget* **12**, 1707–1723 (2021).

225. Tang, W., Duan, J., Zhang, J. G. & Wang, Y. P. Subtyping glioblastoma by combining miRNA and mRNA expression data using compressed sensing-based approach. *Eurasip J. Bioinforma. Syst. Biol.* **2013**, 1–9 (2013).

226. Filbin, M. G. *et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science (80-.).* **360**, 331–335 (2018).

227. Zhang, Z. *et al.* Uniform genomic data analysis in the NCI Genomic Data Commons. *Nat. Commun.* **12**, 1–11 (2021).

228. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

229. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

230. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).

231. Pantano, L., Estivill, X. & Martí, E. SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. *Nucleic Acids Res.* **38**, e34–e34 (2010).

232. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

233. Maciej Pajak, T. I. S. miRNAtap: miRNAtap: microRNA Targets— Aggregated Predictions. (2019).

234. Ahrens, T. D. *et al.* The Role of Proteoglycans in Cancer Metastasis and Circulating Tumor Cell Analysis. *Front. Cell Dev. Biol.* **8**, (2020).

235. Misra, J. R. & Irvine, K. D. The Hippo Signaling Network and Its Biological Functions. *Annu. Rev. Genet.* **52**, 65–87 (2018).

236. Farhan, M. *et al.* FOXO Signaling Pathways as Therapeutic Targets in Cancer. *Int. J. Biol. Sci.* **13**, 815–827 (2017).

237. Guo, Y. *et al.* ERK/MAPK signalling pathway and tumorigenesis (Review). *Exp. Ther. Med.* (2020) doi:10.3892/etm.2020.8454.

238. Patel, S., Alam, A., Pant, R. & Chattopadhyay, S. Wnt Signaling and Its Significance Within the Tumor Microenvironment: Novel Therapeutic Insights. *Front. Immunol.* **10**, (2019).

239. Katoh, Y. & Katoh, M. Hedgehog signaling, epithelial-to-mesenchymal transition and miRNA (review). *Int. J. Mol. Med.* **22**, 271–5 (2008).

240. Yeo, S. K. *et al.* Single-cell RNA-sequencing reveals distinct patterns of cell state heterogeneity in mouse models of breast cancer. *Elife* **9**, 1–24 (2020).

241. Song, H. *et al.* Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nat. Commun.* **13**, (2022).

242. Reczko, M., Maragkakis, M., Alexiou, P., Papadopoulos, G. L. & Hatzigeorgiou, A. G. Accurate microRNA Target Prediction Using Detailed Binding Site Accessibility and Machine Learning on Proteomics Data. *Front. Genet.* **2**, (2012).

243. Agarwal, V., Bell, G. W., Nam, J.-W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, (2015).

244. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37**, 495–500 (2005).

245. Miranda, K. C. *et al.* A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. *Cell* **126**, 1203–1217 (2006).

246. Wong, N. & Wang, X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* **43**, D146–D152 (2015).

247. Rojo Arias, J. & Busskamp, V. Challenges in microRNAs' targetome prediction and validation. *Neural Regen. Res.* **14**, 1672–1677 (2019).

248. Riolo, G., Cantara, S., Marzocchi, C. & Ricci, C. miRNA Targets: From Prediction Tools to Experimental Validation. *Methods Protoc.* **4**, 1 (2020).

249. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654–665 (2013).

250. Wang, X. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. *Bioinformatics* **32**, 1316–1322 (2016).

251. McGeary, S. E. *et al.* The biochemical basis of microRNA targeting efficacy. *Science (80-.).* **366**, (2019).

252. Arora, A. & Simpson, D. A. Individual mRNA expression profiles reveal the effects of specific microRNAs. *Genome Biol.* **9**, R82 (2008).

253. Liu, W. & Wang, X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol.* **20**, 18 (2019).

254. Lähnemann, D. *et al.* Eleven grand challenges in single-cell data science. *Genome Biol.* **21**, 31 (2020).

255. Guo, L. & Chen, F. A challenge for miRNA: Multiple isomiRs in miRNAomics. *Gene* **544**, 1–7 (2014).

256. Neilsen, C. T., Goodall, G. J. & Bracken, C. P. IsomiRs - The overlooked repertoire in the dynamic microRNAome. *Trends Genet.* **28**, 544–549 (2012).

257. Ros, X. B.-D., Yang, A. & Gu, S. IsomiRs: Expanding the miRNA repression toolbox beyond the seed. *Biochim. Biophys. Acta - Gene Regul. Mech.* (2019) doi:10.1016/j.bbagrm.2019.03.005.

258. Hagemann-Jensen, M., Abdullayev, I., Sandberg, R. & Faridani, O. R. Small-seq for single-cell small-RNA sequencing. *Nat. Protoc.* 1–18 doi:10.1038/s41596-018-0049-y.

259. Yu, F. *et al.* Naturally existing isoforms of miR-222 have distinct functions. *Nucleic Acids Res.* **45**, 11371–11385 (2017).

260. Amsel, D., Vilcinskas, A. & Billion, A. Evaluation of high-throughput isomiR identification tools: Illuminating the early isomiRome of Tribolium castaneum. *BMC Bioinformatics* **18**, 1–13 (2017).

261. Fu, Y., Wu, P. H., Beane, T., Zamore, P. D. & Weng, Z. Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics* **19**, 1–14 (2018).

262. Ibuki, Y. *et al.* Circulating microRNA/isomiRs as novel biomarkers of esophageal squamous cell carcinoma. *PLoS One* **15**, 1–18 (2020).

263. Schirmer, M. *et al.* Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43**, e37–e37 (2015).

264. Pfeiffer, F. *et al.* Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci. Rep.* **8**, 1–14 (2018).

265. Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V. & Polz, M. F. PCR-induced sequence artifacts and bias: Insights from comparison of two 16s rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* **71**, 8966–8969 (2005).

266. Daniel, A. *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, 1–14 (2011).

267. Zhang, X., Ping, P., Hutvagner, G., Blumenstein, M. & Li, J. Aberration-corrected ultrafine analysis of miRNA reads at single-base resolution: a k-mer lattice approach. *Nucleic Acids Res.* **49**, e106 (2021).

268. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2012).

269. Kou, R. *et al.* Benefits and challenges with applying unique molecular identifiers in next generation sequencing to detect low frequency mutations. *PLoS One* **11**, 1–15 (2016).

270. Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. The impact of amplification on differential expression analyses by RNA-seq. *Sci. Rep.* **6**, 1–11 (2016).

271. Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **65**, 631-643.e4 (2017).

272. Zhu, T., Liao, K., Zhou, R., Xia, C. & Xie, W. ATAC-seq with unique molecular identifiers improves quantification and footprinting. *Commun. Biol.* **3**, (2020).

273. Saunders, K. *et al.* Insufficiently complex unique-molecular identifiers (UMIs) distort small RNA sequencing. *Sci. Rep.* **10**, 1–9 (2020).

274. Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nat. Genet.* **43**, 854–859 (2011).

275. Gutiérrez-Vázquez, C. *et al.* 3′ Uridylation controls mature microRNA turnover during CD4 T-cell activation. *Rna* **23**, 882–891 (2017).

276. Yang, A. *et al.* 3′ Uridylation Confers miRNAs with Non-canonical Target Repertoires. *Mol. Cell* **75**, 511-522.e4 (2019).

277. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014).

278. Sheu-Gruttadauria, J., Xiao, Y., Gebert, L. F. & MacRae, I. J. Beyond the seed: structural basis for supplementary microRNA targeting by human Argonaute2. *EMBO J.* **38**, 1–14 (2019).

279. Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D. & Serebrov, V. Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell* **162**, 84–95 (2015).

280. Sena, J. A. *et al.* Unique Molecular Identifiers reveal a novel sequencing artefact with implications for RNA-Seq based gene expression analysis. *Sci. Rep.* **8**, 1–13 (2018).

281. Sanchez Herrero, J. F., Pluvinet, R., Luna de Haro, A. & Sumoy, L. Paired-end small RNA sequencing reveals a possible overestimation in the isomiR sequence repertoire previously reported from conventional single read data analysis. *BMC Bioinformatics* **22**, 1–16 (2021).

282. Qiu, P. Embracing the dropouts in single-cell RNA-seq analysis. *Nat. Commun.* **11**, 1–9 (2020).

283. Wright, C. *et al.* Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. *BMC Genomics* **20**, 1–21 (2019).

284. Dard-dascot, C. *et al.* Systematic comparison of small RNA library preparation protocols for next-generation sequencing. 1–16 (2018) doi:10.1186/s12864-018-4491-6.

285. Yeri, A. *et al.* Evaluation of commercially available small RNASeq library preparation kits using low input RNA. *BMC Genomics* **19**, 1–15 (2018).

286. Fuchs, R. T., Sun, Z., Zhuang, F. & Robb, G. B. Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS One* **10**, 1–24 (2015).

287. Mitchell, K. *et al.* Benchmarking of computational error-correction methods for next-generation sequencing data. *Genome Biol.* **21**, 1–13 (2020).

288. Maguire, S., Lohman, G. J. S. & Guan, S. A low-bias and sensitive small RNA library preparation method using randomized splint ligation. *Nucleic Acids Res.* **48**, 1–14 (2020).

289. Jayaprakash, A. D., Jabado, O., Brown, B. D. & Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* **39**, 1–12 (2011).

290. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Front. Endocrinol. (Lausanne).* **9**, 1–12 (2018).

291. Klein, M., Chandradoss, S. D., Depken, M. & Joo, C. Why Argonaute is needed to make microRNA target search fast and reliable. *Semin. Cell Dev. Biol.* **65**, 20–28 (2017).

292. Daugaard, I. & Hansen, T. B. Biogenesis and Function of Ago-Associated RNAs. *Trends Genet.* **33**, 208–219 (2017).

293. Dueck, A., Ziegler, C., Eichner, A., Berezikov, E. & Meister, G. MicroRNAs associated with the different human Argonaute proteins. *Nucleic Acids Res.* **40**, 9850–9862 (2012).

294. Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).

295. Ketting, R. F. & Cochella, L. Concepts and functions of small RNA pathways in C. elegans. in 45–89 (2021). doi:10.1016/bs.ctdb.2020.08.002.

296. Hutvágner, G. & Zamore, P. D. RNAi: nature abhors a double-strand. *Curr. Opin. Genet. Dev.* **12**, 225–232 (2002).

297. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, 501–504 (2005).

298. Harrow, J. *et al.* GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res.* **22**, 1760–1774 (2012).

299. Medley, J. C., Panzade, G. & Zinovyeva, A. Y. microRNA strand selection: Unwinding the rules. *Wiley Interdiscip. Rev. RNA* **12**, 1–22 (2021).

300. Meijer, H. A., Smith, E. M. & Bushell, M. Regulation of miRNA strand selection: Follow the leader? *Biochem. Soc. Trans.* **42**, 1135–1140 (2014).

301. Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).

302. Ro, S., Park, C., Young, D., Sanders, K. M. & Yan, W. Tissue-dependent paired expression of miRNAs. *Nucleic Acids Res.* **35**, 5944–5953 (2007).

303. Hu, H. Y. *et al.* Sequence features associated with microRNA strand selection in humans and flies. *BMC Genomics* **10**, 413 (2009).

304. Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818–822 (2010).

305. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

306. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).

307. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

308. Mercer, T. R. *et al.* Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat. Protoc.* **9**, 989–1009 (2014).

309. Okamura, K., Liu, N. & Lai, E. C. Distinct Mechanisms for MicroRNA Strand Selection by Drosophila Argonautes. *Mol. Cell* **36**, 431–444 (2009).

310. Petri, R. & Jakobsson, J. Identifying miRNA targets using AGO-RIPseq. *Methods Mol. Biol.* **1720**, 131–140 (2018).

311. Kadekar, S. *et al.* Synthetic Design of Asymmetric miRNA with an Engineered 3′ Overhang to Improve Strand Selection. *Mol. Ther. - Nucleic Acids* **16**, 597–604 (2019).

312. Wilson, R. C. & Doudna, J. A. The Dicer-TRBP Interface Structure and Implications for Strand Selection During Microrna Biogenesis. *Biophys. J.* **106**, 463a (2014).

313. Kim, H. *et al.* A Mechanism for microRNA Arm Switching Regulated by Uridylation. *Mol. Cell* **78**, 1224-1236.e5 (2020).

314. Li, L. *et al.* The landscape of miRNA editing in animals and its impact on miRNA biogenesis and targeting. *Genome Res.* **28**, 132–143 (2018).

315. Ui-Tei, K., Nishi, K., Takahashi, T. & Nagasawa, T. Thermodynamic Control of Small RNA-Mediated Gene Silencing. *Front. Genet.* **3**, (2012).

316. Li, Y. *et al.* Let-7b-3p inhibits tumor growth and metastasis by targeting the BRF2-mediated MAPK/ERK pathway in human lung adenocarcinoma. *Transl. Lung Cancer Res.* **10**, 1841–1856 (2021).

317. Bernstein, D. L., Jiang, X. & Rom, S. Let-7 microrna: Their role in cerebral and cardiovascular diseases, inflammation, cancer, and their regulation. *Biomedicines* **9**, 1–18 (2021).

318. Zhang, S. *et al.* Silencing protein kinase C ζ by microRNA-25-5p activates AMPK signaling and inhibits colorectal cancer cell proliferation. *Oncotarget* **8**, 65329–65338 (2017).

319. Wang, Y. *et al.* Melatonin Inhibits the Progression of Oral Squamous Cell Carcinoma via Inducing miR-25-5p Expression by Directly Targeting NEDD9. *Front. Oncol.* **10**, (2020).

320. Liu, B. *et al.* MicroRNA-27b inhibits cell proliferation in oral squamous cell carcinoma by targeting FZD7 and Wnt signaling pathway. *Arch. Oral Biol.* **83**, 92–96 (2017).

321. Liu, C. H. *et al.* Tumor-suppressor miRNA-27b-5p regulates the growth and metastatic behaviors of ovarian carcinoma cells by targeting CXCL1. *J. Ovarian Res.* **13**, 92 (2020).

322. Guo, J. Transcription: the epicenter of gene expression. *J. Zhejiang Univ. Sci. B* **15**, 409–411 (2014).

323. Gibney, E. R. & Nolan, C. M. Epigenetics and gene expression. *Heredity (Edinb).* **105**, 4–13 (2010).

324. Li, W., Li, F., Zhang, X., Lin, H.-K. & Xu, C. Insights into the post-translational modification and its emerging role in shaping the tumor microenvironment. *Signal Transduct. Target. Ther.* **6**, 422 (2021).

325. Wilczynska, A. & Bushell, M. The complexity of miRNA-mediated repression. *Cell Death Differ.* **22**, 22–33 (2015).

326. Enfield, K. S. S. *et al.* Deregulation of small non-coding RNAs at the DLK1-DIO3 imprinted locus predicts lung cancer patient outcome. *Oncotarget* **7**, 80957–80966 (2016).

327. LaBelle, J. *et al.* Commercially Available Blocking Oligonucleotides Effectively Suppress Unwanted Hemolysis-Related miRNAs in a Large Whole-Blood RNA Cohort. *J. Mol. Diagnostics* **23**, 671–682 (2021).

328. Elhamamsy, A. R. *et al.* Circulating miR-92a, miR-143 and miR-342 in Plasma are Novel Potential Biomarkers for Acute Myeloid Leukemia. *Int. J. Mol. Cell. Med.* **6**, 77–86 (2017).

329. Li, L.-M., Wang, D. & Zen, K. MicroRNAs in Drug-induced Liver Injury. *J. Clin. Transl. Hepatol.* **2**, 162–9 (2014).

330. de Oliveira, L. F. *et al.* Differential expression analysis and profiling of hepatic miRNA and isomiRNA in dengue hemorrhagic fever. *Sci. Rep.* **11**, 5554 (2021).

331. Papageorgiou, S. G. *et al.* MicroRNA-92a-3p overexpression in peripheral blood mononuclear cells is an independent predictor of prolonged overall survival of patients with chronic lymphocytic leukemia. *Leuk. Lymphoma* **60**, 658–667 (2019).

332. Zhang, Y., Qian, J., Gu, C. & Yang, Y. Alternative splicing and cancer: a systematic review. *Signal Transduct. Target. Ther.* **6**, 78 (2021).

333. Bonnal, S. C., López-Oreja, I. & Valcárcel, J. Roles and mechanisms of alternative splicing in cancer — implications for care. *Nat. Rev. Clin. Oncol.* **17**, 457–474 (2020).

334. Boyerinas, B., Park, S.-M., Hau, A., Murmann, A. E. & Peter, M. E. The role of let-7 in cell differentiation and cancer. *Endocr. Relat. Cancer* **17**, F19–F36 (2010).

335. Caiazza, C. & Mallardo, M. The Roles of miR-25 and its Targeted Genes in Development of Human Cancer. *MicroRNA (Shariqah, United Arab Emirates)* **5**, 113–119 (2016).

336. Zhang, J. *et al.* MiR-27b suppresses epithelial–mesenchymal transition and chemoresistance in lung cancer by targeting Snail1. *Life Sci.* **254**, 117238 (2020).

337. Peng, Y. & Croce, C. M. The role of microRNAs in human cancer. *Signal Transduct. Target. Ther.* **1**, (2016).

338. Chatterjee, N., Rana, S., Espinosa-Diez, C. & Anand, S. MicroRNAs in Cancer: Challenges and Opportunities in Early Detection, Disease Monitoring, and Therapeutic Agents. *Curr. Pathobiol. Rep.* **5**, 35–42 (2017).

339. Forterre, A., Komuro, H., Aminova, S. & Harada, M. A Comprehensive Review of Cancer MicroRNA Therapeutic Delivery Strategies. *Cancers (Basel).* **12**, 1852 (2020).

340. Bajan, S. & Hutvagner, G. RNA-Based Therapeutics: From Antisense Oligonucleotides to miRNAs. *Cells* **9**, 137 (2020).

341. Li, Y. *et al.* A comprehensive genomic pan-cancer classification using The Cancer Genome Atlas gene expression data. *BMC Genomics* **18**, 508 (2017).

342. Zadran, S., Remacle, F. & Levine, R. D. miRNA and mRNA cancer signatures determined by analysis of expression levels in large cohorts of patients. *Proc. Natl. Acad. Sci.* **110**, 19160–19165 (2013).

343. Tsang, E. K. *et al.* Small RNA sequencing in cells and exosomes identifies eQTLs and 14q32 as a region of active export. *G3 Genes, Genomes, Genet.* **7**, 31–39 (2017).