

**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 heterogeneous 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

Abstract.....	15
1. Literature review.....	16
1.1. miRNA biogenesis	16
1.1.1. miRNA biogenesis via the canonical pathway	16
1.1.2. Non-canonical pathways for miRNA biogenesis.....	17
1.1.3. Alternative splicing regulates the function of miRNA biogenesis proteins.....	18
1.1.4. miRNAs often form clusters.....	18
1.2. IsomiRs: a hidden layer of gene regulation	21
1.2.1. Uridylation regulates the processing of several pre-miRNAs.....	21
1.2.2. IsomiRs with distinct regulatory functions	21
1.2.3. miRNA stability and turnover is regulated by 3' tailing.....	22
1.3. Argonaute is essential for miRNA-mediated translation repression.....	24
1.3.1. Functional domains of Argonaute	24
1.3.2. Mechanisms of translational repression by the RISC	27
1.4. miRNA dysregulation is a general feature in many cancers	29
1.4.1. Impairments of the miRNA processing machinery in cancer	29
1.4.2. Genetic aberrations of miRNA genes in cancer	31
1.4.3. Epigenetic regulation of miRNAs is affected in cancer	32
1.4.4. miRNAs are key mediators of transcription factor-driven oncogenesis.....	32
1.4.5. IsomiR expression has been linked to cancer	35
1.4.6. Non-coding RNAs in pediatric solid tumours.....	35
1.5. Single cell sequencing reveals genetic and transcriptomic heterogeneity in cancers .	36
1.5.1. Single cell sequencing has enabled deeper interrogation of intra-tumour heterogeneity.....	37
1.5.2. The current state of single cell small RNA sequencing.....	38
1.6. Summary and research questions	41
1.6.1. Hypothesis 1 - microRNA and isomiR expression contributes to heterogeneous gene regulation and can identify subpopulations in cancer	42
1.6.2. Hypothesis 2 - Splice variants of Argonaute 2 have distinct regulatory functions through association with different RNAs.....	42
2. The role of miRNAs in glioblastoma tumour heterogeneity and cell state regulation.....	43
2.1. Chapter introduction	43
2.1.1. Heterogeneity is a hallmark feature of the brain cancer glioblastoma.....	43

2.1.2. Evolution of models for intra-tumour heterogeneity in glioblastoma	43
2.1.3. miRNAs play a key role in brain development and glioblastoma	44
2.2. Specific methodology.....	45
2.2.1. Data collection	45
2.2.2. Single cell small RNA sequencing pre-processing and mapping.....	45
2.2.3. Identification of cell subpopulations with miRNA expression.....	45
2.2.4. miRNA target prediction	46
2.2.5. Scoring for TCGA subtypes, cell states and miRNA targets	46
2.2.6. Code for data analysis and figures.....	46
2.3. Identification of heterogeneously expressed miRNAs in glioblastoma.....	47
2.3.1. Distinct glioblastoma subpopulations can be identified through single cell miRNA and isomiR expression profiles	47
2.3.2. Expression of two miRNA clusters distinguish KS4 glioblastoma subpopulations	49
2.4. Cell autonomous miRNAs are associated with glioblastoma subtypes and cell state genes.....	54
2.4.1. Different miRNAs are associated with TCGA glioblastoma subtypes.....	54
2.4.2. Different miRNAs are associated with glioblastoma cell state genes	58
2.4.3. Two miRNA clusters are associated with different cell states in glioblastoma cells	60
2.5. The DLk1-Dio3 locus is associated with cell states in single cells.....	62
2.5.1. Long non-coding RNAs MEG3, MEG8 and MEG9 predict Dlk1-Dio3 miRNA expression in glioblastoma tumours.....	62
2.5.2. MEG3 expression is associated with the neural progenitor-like and oligodendrocyte progenitor-like cell states in single cells.....	64
2.6. Target prediction identifies Dlk1-Dio3 miRNAs regulate key transcription factors and regulatory pathways involved in cancer	67
2.6.2. Expression of miRNA targets is associated positively with NPC and OPC cell states and negatively with MES states	71
2.7. Discussion.....	79
2.7.1. Defining glioblastoma cell states	79
2.7.2. Challenges of determining miRNA activity through changes in target expression	80
2.7.3. miRNAs are potential biomarkers of glioblastoma heterogeneity.....	81
3. A comparative analysis of single cell small RNA sequencing data reveals heterogenous isomiR expression and regulation.....	83

3.1. Chapter introduction	83
3.2. Specific methodology.....	85
3.2.1. Data Collection.....	85
3.2.2. Sequencing Data Processing	85
3.2.3. miRNA Mapping and Annotation.....	85
3.2.4. Correlation of miRNA and isomiR Expression with Predicted Targets	86
3.2.5. Code for Data Analysis and Figures	86
3.3. miRNA and IsomiR abundance are highly variable across cell types in the three single cell small RNA-seq protocols	87
3.4. IsomiR processing is likely regulated by cell autonomous mechanisms	95
3.4.1. Global expression of isomiRs is specific to cell types	95
3.4.2. Processing of isomiRs is unique to each miRNA's gene.....	96
3.4.3. Individual cells exhibit cell type specific isomiR processing	96
3.5. IsomiR processing alters regulation of their canonical targets	101
3.6. Unique Molecular Identifiers increase estimated isomiR abundance	105
3.7. Discussion.....	110
3.7.1. Analysis of isomiR types overcomes current limitations with single cell small RNA sequencing data	110
3.7.2. Biases in small RNA sequencing protocols are likely to influence isomiR quantification.....	111
3.7.3. Methods to improve single cell small RNA sequencing data quality for isomiR analysis.....	112
3.7.4. Implications of cell autonomous isomiR regulation in cancer.....	112
4. Investigating the RNA binding properties of alternatively spliced Argonaute 2 (Ago2) isoforms	113
4.1. Chapter introduction	113
4.1.1. Argonaute loads multiple classes of small RNAs	113
4.1.2. Features of the miRNA duplex influence Argonaute strand selection	114
4.1.3. Alternative splicing may produce Argonaute proteins with altered functions ...	115
4.2. Specific methodology.....	116
4.2.1. Buffers.....	116
4.2.2. Reagents.....	116
4.2.3. Antibodies	116
4.2.4. Cell provision.....	117
4.2.5. Ago2 isoform plasmid transfection.....	117

4.2.6. Cell lysis for immunoprecipitation	117
4.2.7. Immunoprecipitation of FLAG-tagged Ago2 isoforms	117
4.2.8. RNA extraction	118
4.2.9. Sequencing data generation	118
4.2.10. Calculation of Percent Spliced with Junction (PSJ) values	118
4.2.11. Small RNA biotype quantification	119
4.2.12. miRNA and isomiR quantification	119
4.2.13. Differential expression analysis	119
4.2.14. Estimating base preferences in Ago2 isoforms	120
4.2.15. Code for Data Analysis and Figures	120
4.3. Alternative Splicing of Ago2 Suggests Diversity of Isoform Expression in Human Tissues	121
4.4. Alternatively spliced Ago2 isoforms do not bind miRNAs as strongly as the canonical isoform	124
4.5. Ago2 isoforms have altered miRNA binding preferences.....	127
4.5.1. The relative abundance of certain miRNAs is altered in Ago2 isoforms	127
4.5.2. Alternative splicing of Ago2 changes miRNA strand selection preference	127
4.6. The 5' nucleotide preference during strand selection is altered in Ago2 isoforms ...	132
4.7. Ago2 isoforms have altered isomiR associations.....	136
4.8. Discussion.....	144
4.8.1. Limitations of sequencing data for the investigation of Argonaute strand selection	144
4.8.2. Ago2 isoforms may interact with isomiRs to fine-tune miRNA strand selection	145
4.8.3. Implications of Argonaute isoforms altering miRNA activity in normal and cancerous cells	145
5. Final discussion	147
5.1. A multitude of mechanisms regulate the miRNA pathway	147
5.2. Different mechanisms of miRNA regulation may be perturbed in cancers	148
5.3. miRNAs as biomarkers in cancer	149
5.4. Concluding Remarks.....	150
Appendix A - Non-Coding RNAs in Pediatric Solid Tumors	151
Appendix B – Statistical Values and Read Counts for IsomiR-Target Pairs.....	170
Appendix C – Read Counts for Spliced Exon Junctions in Ago2.....	174
Appendix D – Western Blots for the Immunoprecipitation of Ago2 Isoforms	178

Bibliography179

List of Figures and Tables

1. Literature Review

Figures	Page
Figure 1.1. Pathways of miRNA biogenesis	20
Figure 1.2. miRNAs are commonly expressed as various isoforms called isomiRs	23
Figure 1.3. Argonaute consists of four primary functional domains	26
Figure 1.4. Multiple mechanisms facilitate post-transcriptional gene regulation with the RISC	28
Figure 1.5. Gene regulation involves complex interactions between miRNAs, transcription factors, and long non-coding RNAs which all be dysregulated in cancers	34
Figure 1.6. Molecular heterogeneity is evident in genomic, epigenetic, transcriptomic, and proteomic profiles	40

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

Figures	Page
Figure 2.1. Identification of heterogenous glioblastoma cell subpopulations from miRNA and isomiR expression	48
Figure 2.2. Comparison of miRNAs between KS4 glioblastoma cells in miRNA-based clusters 1 (red) and 2 (blue)	50
Figure 2.3. DLK1-DIO3 locus derived miRNAs upregulated in KS4 Glioblastoma miRNA-based clusters 1 (red) and 2 (blue)	51
Figure 2.4. Expression of upregulated miRNAs in two KS4 cell subpopulations is dominated by two miRNA clusters	52
Figure 2.5. Comparison of isomiRs between KS4 Glioblastoma cells in isomiR-based clusters 1 (red) and 2 (blue)	53
Figure 2.6. Comparison of each pair of TCGA subtype scores in glioblastoma tumours	55
Figure 2.7. Comparison of TCGA subtype and cell state glioblastoma gene modules and their association with miRNAs	56
Figure 2.8. Expression of miRNAs correlate with different TCGA subtypes	57
Figure 2.9. Expression of miRNAs correlate with different cell state modules	59
Figure 2.10. Expression of Dlk1-Dio3 and miR-224/452 cluster miRNAs correlate with different cell state modules	61

Figure 2.11. Correlation of MEG3 expression with cell state module scores across all glioblastoma cells for each dataset	65
Figure 2.12. MEG3 expression and correlation with cell state module scores in cells from each glioblastoma tumour	66
Figure 2.13. Common genes between cell state modules and miRNA targets	69
Figure 2.14. Gene Set Enrichment Analysis of KEGG pathways for miRNA cluster targets	70
Figure 2.15. Expression of gene targets for miRNAs from the Dlk1-Dio3 cluster	73
Figure 2.16. Expression of gene targets for miRNAs from the miR-224/452 cluster	74
Figure 2.17. Cluster target scores for the Dlk1-Dio3 cluster	75
Figure 2.18. Cluster target scores for the miR-224/452 cluster	76
Figure 2.19. Correlation of scores for cluster targets and cell states in glioblastoma tumours	77
Figure 2.20. Correlation of miRNA target scores for 200 miRNAs highly expressed in glioblastoma tumours	78

Tables	Page
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	63

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

Figures	Page
Figure 3.1. An overview of the small RNA sequencing samples that were analysed	89
Figure 3.2. Top 20 miRNAs by expression for each cell type included in this study	91
Figure 3.3. Comparison of total miRNA and isomiR length distribution profiles between single cell and bulk RNA-seq datasets from independent studies	92
Figure 3.4. Expression and relative abundance of isomiRs in Hepatocellular Carcinoma single cell and bulk smRNA-seq data from independent studies	93
Figure 3.5. Expression and relative abundance of isomiRs in K562 Leukemia single cell and bulk smRNA-seq data from independent studies	94
Figure 3.6. Comparison of relative isomiR expression in each cell type	98

Figure 3.7. Comparison of 5' and 3' isomiRs in select miRNAs that are highly expressed across multiple cell types	99
Figure 3.8. Expression and relative abundance of isomiRs in single cells	100
Figure 3.9. Correlation of expression levels between canonical miRNAs or isomiR categories and the canonical miRNAs predicted targets (mRNAs) in K562 Leukemia cells	103
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	104
Figure 3.11. Impact of UMI deduplication on isomiR types	106
Figure 3.12. Impact of UMI deduplication on miRNA length distributions in glioblastoma cell lines	107
Figure 3.13. Impact of UMI deduplication on miRNA length distributions in the HEK293 cell line, and naïve and primed embryonic stem cells	108
Figure 3.14. Comparison of 5' and 3' modifications between averaged non-deduplicated (-) and deduplicated (+) glioblastoma cells, for selected miRNAs	109

Tables	Page
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	90

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

Figures	Page
Figure 4.1. Alternative splicing events from RNA Capture-seq read data indicate multiple isoforms of Ago2 are expressed in humans	123
Figure 4.2. Proportion of small RNA classes immunoprecipitated from Ago2 isoforms in HeLa cells	126
Figure 4.3. Comparison of miRNA expression between isoforms	129
Figure 4.4. Isoforms show strand selection bias in specific miRNAs	131
Figure 4.5. Base preferences from all miRNA duplexes for each pair of bases	134
Figure 4.6. Base preferences of miRNA duplexes from each precursor miRNA	135
Figure 4.7. Comparison of the size of miRNAs associated with the three Ago2 isoforms	138

Figure 4.8. IsomiR types associated with Ago2 isoforms	139
Figure 4.9. Proportion of miRNAs associated with Ago2 isoforms, annotated as isomiRs (non-canonical miRNAs) and substitution variants	140
Figure 4.10. miRNAs associated with Ago2 isoforms annotated as 5' variants	141
Figure 4.11. miRNAs associated with Ago2 isoforms annotated as 3' templated variants	142
Figure 4.12. miRNAs associated with Ago2 isoforms annotated as 3' non-templated variants	143

Tables	Page
Table 4.1. Antibodies used for immunoprecipitation experiments	116
Table 4.2. Differentially expressed miRNAs when comparing three isoforms of Ago2	130

Appendix

Figures	Page
Figure Appendix D. Western blots confirm enrichment of FLAG antibodies in FLAG immunoprecipitate compared to control for all three Ago2 isoforms	178

Tables	Page
Table Appendix B1. Statistical Values and Read Counts for IsomiR-Target Pairs in K562 cells	170
Table Appendix B2. Statistical Values and Read Counts for IsomiR-Target Pairs in Hepatocellular Carcinoma cells	172
Table Appendix C. Read Counts and Percent Spliced with Junction (PSJ) values for Spliced Exon Junctions in Ago2	174

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.