



OPEN

MiRNAs from the Dlk1-Dio3 locus and miR-224/452 cluster contribute to glioblastoma tumor heterogeneity

Christopher M. Smith^{1,2}, Daniel Catchpole^{3,4} & Gyorgy Hutvagner¹✉

Glioblastoma is one of the most common and aggressive brain tumors and has seen few improvements in patient outcomes. Inter-tumor heterogeneity between tumors of different patients as well as intra-tumor heterogeneity of cells within the same tumor challenge the development of effective drugs. MiRNAs play an essential role throughout the developing brain and regulate many key genes involved in oncogenesis, yet their role in driving many of the processes underlying tumor heterogeneity remains unclear. In this study, we highlight miRNAs from the Dlk1-Dio3 and miR-224/452 clusters which may be expressed cell autonomously and have expression that is associated with cell state genes in glioblastoma, most prominently in neural progenitor-like and mesenchymal-like states respectively. These findings implicate these miRNA clusters as potential regulators of glioblastoma intra-tumoral heterogeneity and may serve as valuable biomarkers for cell state identification.

Glioblastoma is one of the most common and aggressive brain tumors and has seen few improvements in patient outcomes¹. Inter-tumor heterogeneity between tumors of different patients as well as intra-tumor heterogeneity of cells within the same tumor challenge the development of effective drugs². In glioblastoma, inter-tumor heterogeneity has been well characterized through large scale studies integrating gene expression data from The Cancer Genome Atlas (TCGA), initially identifying at least four key subtypes of glioblastoma named proneural (TCGA-PN), neural (TCGA-NE), classical (TCGA-CL), and mesenchymal (TCGA-MS), later revised to exclude the neural subtype^{3,4}. More recently, intra-tumor heterogeneity in glioblastoma was characterized using single cell RNA-seq, which has provided a unified model for glioblastoma heterogeneity⁵. In this model they described four main cancer subpopulations, referred to as cell states, resembling cell types that exist during normal brain development. This included astrocyte-like (AC), neural-progenitor-like (NPC), oligodendrocyte-progenitor-like (OPC), and mesenchymal-like (MES) cell states. Furthermore, they highlighted that tumors were often comprised of multiple cell states, although not necessarily containing all states at once, and that the presence and relative frequencies of each cell state directly influenced TCGA subtype classifications at the population level.

MiRNAs play an essential role throughout the developing brain, contributing to cell fate specification and differentiation in many neural or glial stem/progenitor cells⁶. Extensive documentation of miRNA dysregulation in glioblastoma and their effects on key cancer pathways suggests they are also important in tumorigenesis⁷. Furthermore, miRNA expression profiles can significantly improve classification of tumors with TCGA subtypes, making them potential biomarkers and suggesting their activity is intrinsically linked to gene networks that drive each of these subtypes⁸. 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 regulation⁵. 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. In this study, we highlight that miRNAs from the Dlk1-Dio3 and miR-224/452 clusters may be expressed cell autonomously and that their expression is associated with cell state genes in glioblastoma.

¹School of Biomedical Engineering, Faculty of Engineering and IT, University of Technology Sydney, Sydney, Australia. ²Children's Cancer Institute, Lowy Cancer Research Centre, UNSW Sydney, Sydney, NSW, Australia. ³School of Computer Sciences, Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW, Australia. ⁴The Tumour Bank, The Children's Cancer Research Unit, Kids Research, The Children's Hospital at Westmead, Westmead, NSW, Australia. ✉email: gyorgy.hutvagner@uts.edu.au

Methods

Single cell small RNA sequencing pre-processing and mapping

For the single cell small RNA sequencing reads, UMI sequences were removed prior to 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 sequences⁹. 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 -best¹⁰. Human-aligned reads were subsequently deduplicated with umitools (v1.0.0) with default settings¹¹.

Processed reads were aligned to miRbase (v22.1) annotated precursor miRNAs using miraligner (v3.4), with the following parameters: -sub 1 -trim 3 -add 3¹². Reads which successfully aligned to a miRNA were also annotated with any variants to the miRbase defined mature sequence and converted to miRNA counts.

Identification of cell subpopulations with miRNA expression

To identify glioblastoma cell subpopulations, miRNA expression was analyzed with the R package Seurat (v3.1.5)¹³. MiRNAs present in three or less cells, or cells with less than 1000 mapped reads, were excluded from analysis. For dimensionality reduction, Principal Component Analysis was used on the top 25% of variable features after centering and scaling the data. Cells were visualized using Uniform Manifold Approximation and Projection (UMAP). This was followed by cell clustering using Seurat's FindNeighbors (k.param = 30) and FindClusters (resolution = 0.8) functions. Clustering of cells was stable across a range of parameters and was deemed acceptable due to the presence of cluster specific miRNA 'markers' highly expressed in one population and absent in the other. Seurat's FindAllMarkers function was used for differential expression analysis, and only features with an adjusted *p*-value (*p*.adj) less than 0.05 and log₂ fold-change larger than 1 were considered differentially expressed.

Scoring for TCGA subtypes and cell states

Scores which reflect the expression of a set of genes (i.e. gene modules) were calculated using a method similar to that described in the Neftel et al. study⁵. Briefly, gene counts for all samples were converted to log₂ transcripts-per-million (log₂[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 organize 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 three TCGA subtypes were obtained from Wang et al.'s study⁴. Cell state gene sets were obtained from Neftel et al.'s study⁵.

Results

Expression of the Dlk1-Dio3 and miR-224/452 clusters characterize subpopulations of a glioblastoma primary culture.

To investigate if miRNA heterogeneity may contribute to cell state regulation or identity in glioblastoma, we analyzed publicly available single cell small RNA-seq data for 173 cells from 3 glioblastoma primary cultures and 1 glioblastoma cell line¹⁴. UMAP projections for miRNA expression (Fig. 1A) revealed that most cells colocalized with others from the same cell type with minimal evidence of common miRNA expression modules across different cell types. However, we observed that one of the glioblastoma primary cultures (KS4) formed two distinct groups, not associated with their batch number, that we hypothesized represent two subpopulations in different cell states (Fig. 1B and Supplementary Table 1).

Following this, we compared miRNA expression between the two KS4 cell clusters using differential expression analysis (Fig. 1C–D). 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). Upon further investigation into the miRNAs upregulated in KS4 miRNA cluster 1, we noted that 18 of the 20 upregulated miRNAs originated from the same miRNA cluster hosted within the Dlk1-Dio3 gene locus on chromosome 14q32 (Fig. 1C and Supplementary Fig. 1). Three of the five upregulated miRNAs in cluster 2—miR-224-5p, miR-224-3p, miR-452-5p, were also from a single miRNA cluster on chromosome X (Fig. 1D). Out of the differentially expressed miRNAs identified between the two KS4 subpopulations, expression was dominated (> 75%) by the miRNAs from the Dlk1-Dio3 and miR-224/452 clusters (Fig. 1E–F).

MiRNAs are associated with glioblastoma subtypes and cell state genes.

To predict the function of the upregulated miRNAs in the two KS4 clusters, we analyzed glioblastoma bulk 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. Previous work demonstrated that glioblastoma subtypes are indicative of the most dominant cell states within a tumor, 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) state⁵. 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.

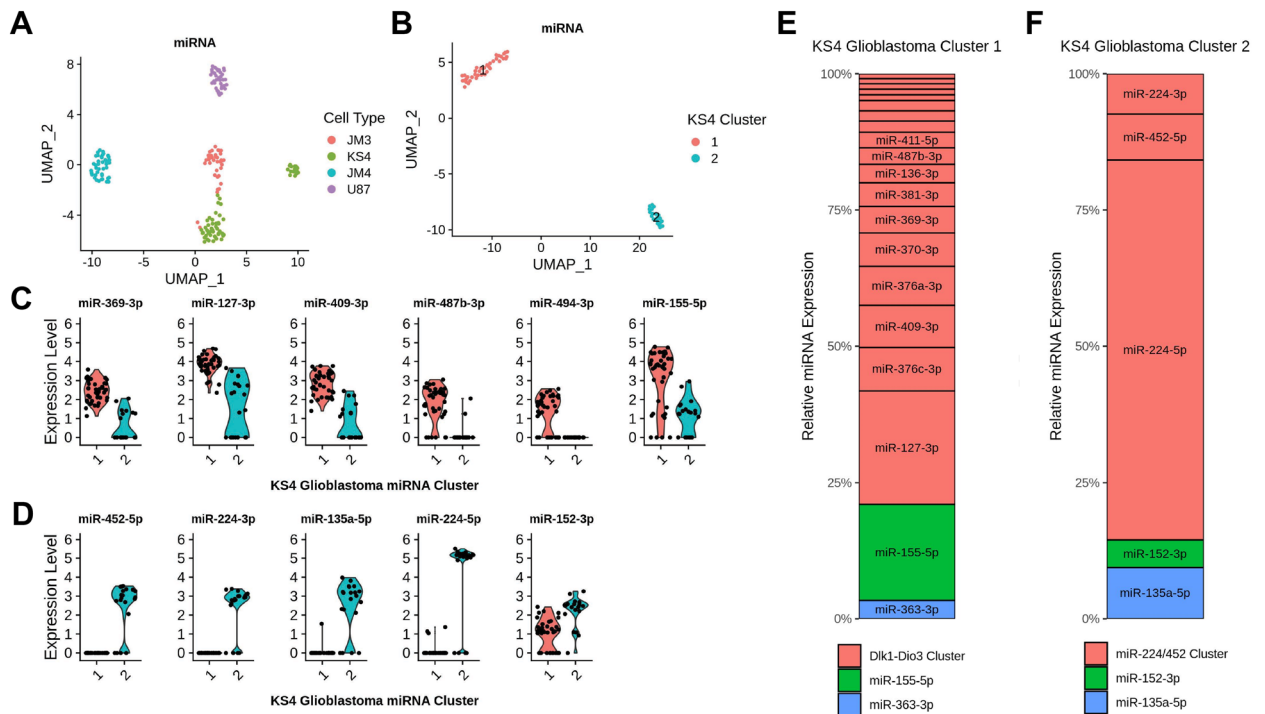


Figure 1. Identification of heterogeneous glioblastoma cell subpopulations from miRNA expression. **(A)** UMAP plot of miRNA expression from four glioblastoma cell types. **(B)** UMAP plot for KS4 glioblastoma cells colored by cluster. Comparison of expression in select miRNAs that were differentially upregulated (adjusted p -value < 0.05) in KS4 cluster 1 **(C)** or KS4 cluster 2 **(D)**. miRNA expression shown is in \log_e . Relative expression of the differentially expressed miRNAs for miRNAs upregulated in **(E)** KS4 Cluster 1 or **(F)** KS4 Cluster 2. miRNAs belonging to the same miRNA cluster were colored similarly.

To detect potential associations between miRNAs and TCGA subtypes or cell states we scored tumors by their mean expression of marker genes associated with subtype or cell state⁵. Tumors were typically associated with at most one TCGA glioblastoma subtype and scores for each TCGA subtype were negatively correlated (Pearson's $R = -0.47$ to -0.31 ; Supplementary Fig. 2). Out of the top 200 expressed miRNAs measured across all glioblastoma tumors, 146 miRNAs were significantly correlated with at least one subtype score ($p_{adj} < 0.05$; Fig. 2A). The strongest correlations of miRNA expression and subtype scores were with the TCGA-PN and TCGA-MS subtypes.

Cell states were defined according to the four cell states described by Neftel et al., and included six gene modules in total, one module associated with each of the AC and OPC states and two modules associated with each of the NPC and MES states⁵. Cell state module scores were correlated against expression of each miRNA (Fig. 2B). The highest Pearson correlation values were observed with the NPC1, NPC2, MES1, and MES2 cell states. Strong anti-correlation patterns with similar magnitudes were observed between the NPC1/NPC2 states and MES1/MES2 states across most miRNAs. A high number of miRNAs had matching correlation signs between NPC1, NPC2, OPC, and AC cell states, although correlation was generally weaker with the OPC and AC cell states.

The correlation of scores and miRNA expression was similar between the TCGA-Proneural subtype and NPC1/NPC2 states, as well as the TCGA-MS subtype and MES1/MES2 states (Fig. 2C). This was consistent with Neftel et al.'s previous work suggesting these subtypes were predominantly composed of cells exhibiting these respective states⁵. 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 subtype's mixed population of states.

The Dlk1-Dio3 and miR-224/452 clusters are associated with glioblastoma cell states.

Focusing on the differentially expressed miRNAs identified between the two KS4 glioblastoma subpopulations, most miRNAs that were co-upregulated in one of these subpopulations had matching signs of correlation coefficients for each cell state (Fig. 2D). For example, 16/20 ($p_{adj} < 0.05$) miRNAs upregulated in KS4 cluster 1 were positively correlated with the NPC1 cell state score, 16/20 ($p_{adj} < 0.05$) with the NPC2 score, and 10/20 ($p_{adj} < 0.05$) with the OPC score. Conversely there was a negative correlation with 8/20 ($p_{adj} < 0.05$) 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 some were weakly correlated or not statistically significant. 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 five miRNAs upregulated in KS4 cluster 2 (Fig. 2D), three were positively correlated with the MES1 scores and two with the MES2 scores, as well as two negatively correlated with NPC1 scores, and one with NPC2 and OPC scores. All three miRNAs from the

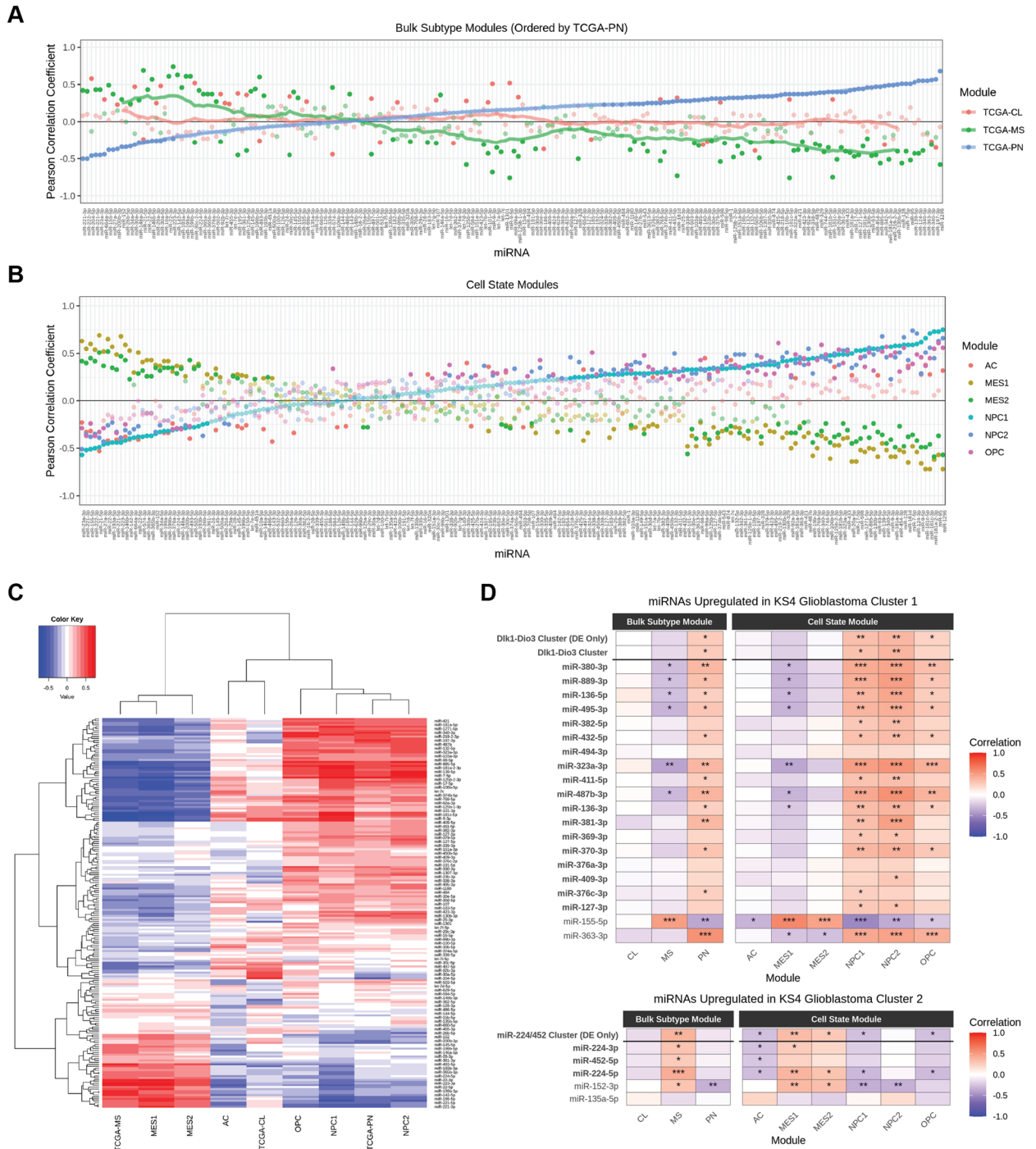


Figure 2. Expression of miRNAs correlate with TCGA subtypes and cell state modules. Pearson correlation for the 200 highest expressed miRNAs and (A) TCGA subtype or (B) cell state module scores. miRNAs are ordered by ascending Pearson correlation coefficients for the Proneural (TCGA-PN) subtype. Non-significant Pearson coefficients are shown as transparent dots. (C) Hierarchical clustering of Pearson correlation values between each miRNA's expression and corresponding gene module score. (D) Pearson correlation of miRNAs upregulated in KS4 cluster 1 (top) or (bottom) KS4 cluster 2 with each TCGA subtype or cell state score. miRNAs belonging to the Dlk1-Dio3 miRNA cluster or miR-224/452 cluster are shown in bold, as well as combined expression of all miRNAs from each cluster (All) or only miRNAs differentially expressed between KS4 cells (DE Only). P-values are shown as asterisks—*: 0.005 < = p-value < 0.05; **: 0.0005 < = p-value < 0.005; ***: 0.00005 < = p-value < 0.0005. CL: Classical; MS: Mesenchymal, PN: Proneural; AC: Astrocyte-like; MES: Mesenchymal-like; NPC: Neural-progenitor-like; OPC: Oligodendrocyte-progenitor-like.

miR-224/452 cluster negatively correlated with AC cell state scores. We also considered that an aggregated expression of miRNAs from each cluster (Fig. 2D), 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 expression (Fig. 2D), correlation 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, 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 (Fig. 2D), 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-tumor heterogeneity in glioblastoma and implicates them as potential regulators of glioblastoma cell states.

The Dlk1-Dio3 cluster lncRNA MEG3 is associated with the neural progenitor-like and oligodendrocyte progenitor-like cell states in single cells.

Previous studies suggest that miRNAs from the Dlk1-Dio3 locus are co-transcribed from a common primary transcript with 3 long non-coding RNAs—MEG3, MEG8, and MEG9¹⁵. To investigate expression of this locus in single cells we analyzed single cell RNA-seq data from four studies. 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.

Firstly, 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 bulk RNA and miRNA expression data for glioblastoma tumors from the CPTAC dataset to determine the correlation of an aggregated 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 (Supplementary Table 2). Excluding genes which code for miRNAs or snoRNAs, these four were the most positively correlated across the glioblastoma tumors, 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.

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 50 glioblastomas (45 unique tumors) and 16,269 cells after filtering^{5,16–18}. Of the Dlk1-Dio3 markers, only MEG3 was detectable in a high number of cells so we focused on this gene.

Glioblastoma cells were assigned a score for each cell state using the same methodology described previously. We calculated the Pearson correlation of cell state module scores and MEG3 expression ($\log_2[\text{TPM} + 1]$) for all cells in each dataset. In most tumors (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; Fig. 3). Tumors with a positive correlation of MEG3 expression and NPC scores generally had a corresponding negative correlation with the MES1 and/or MES2 scores. Although correlations between MEG3 and cell state scores were typically

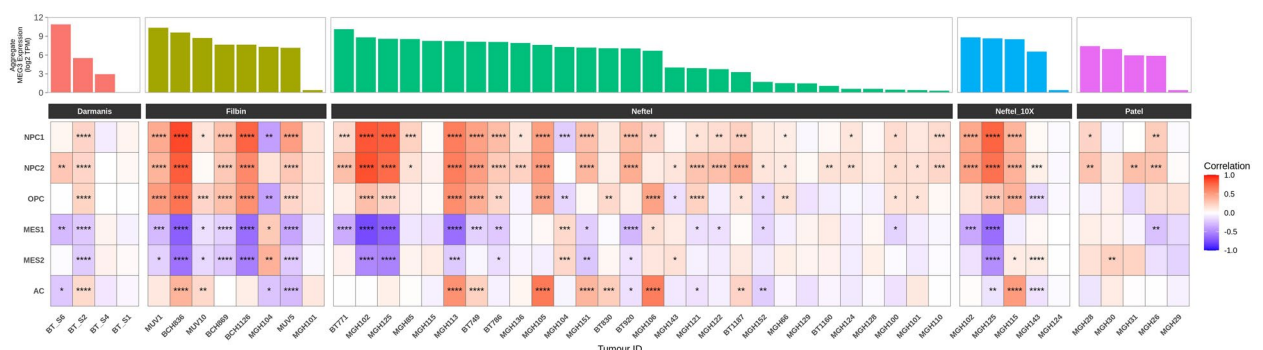


Figure 3. MEG3 is associated with glioblastoma cell states. MEG3 expression (top) and Pearson correlation with cell state module scores (bottom) in glioblastoma cells from different patients. P-value significance is shown as asterisks. *: $0.005 < p\text{-value} < 0.05$; **: $0.0005 < p\text{-value} < 0.005$; ***: $0.00005 < p\text{-value} < 0.0005$; ****: $0.000005 < p\text{-value} < 0.00005$. AC: Astrocyte-like; MES: Mesenchymal-like; NPC: Neural-progenitor-like; OPC: Oligodendrocyte-progenitor-like.

weak to moderate in strength ($R < 0.7$), the results were consistent with our previous findings for the Dlk1-Dio3 miRNAs in the population-level data.

Discussion

In this study we identified miRNAs from the Dlk1-Dio3 and miR-224/452 clusters which distinguished two subpopulations of cells in a primary glioblastoma culture and show that both miRNA clusters are associated with different glioblastoma cell states. Our results suggest miRNA expression from the Dlk1-Dio3 locus, as well as the putatively co-expressed long non-coding RNA MEG3, is positively associated with the NPC cell state and negatively associated with the MES state. The opposite was true with the miR-224/452 cluster where it was positively associated with the MES cell state but negatively associated with the NPC cell state. However, 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, the miRNA highlighted in this study may hold value as biomarkers for cell states once the field evolves towards clinical strategies which factor intra-tumoral heterogeneity.

Recently, a lot of attention has focused on accurately classifying glioblastoma subtypes using smaller gene panels which may be more feasible for use in a clinical setting. The initial Verhaark classification identified a total gene set of 840 genes for classifying classical, mesenchymal, neural, proneural signatures, later reduced to a 48 gene classifier that retained 90.91% accuracy^{3,19}. A 50 gene signature was later defined for IDH-wild type glioblastoma which excluded the neural subtype⁴. A later study by Fu et al. demonstrated 83.7% accuracy in classifying tumors into proneural or mesenchymal subtypes utilizing a smaller set of 26 genes²⁰. None of these studies utilized miRNA expression for classification and our results suggest that a combination of both RNA and miRNA expression may further enhance this accuracy.

A key question is whether defining cell states with RNA or miRNA expression will identify the same cell states or if examining each expression modality will lead to the detection of different subpopulations. In a study with hepatocellular carcinoma cells, the authors used hierarchical clustering to identify three subpopulations of cells with RNA expression but found only two with miRNA expression²¹. There was significant overlap between two 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. However, our study indicates RNA and miRNA expression may classify similar populations.

MiRNAs such as miR-9-3p, miR-27a, and miR-23a, have previously been identified as markers for discriminating TCGA-PN and TCGA-MS subtypes²². Results from our study were consistent with this work, as these three miRNAs had some of the strongest correlation coefficients with these subtypes compared to other miRNAs, supporting the validity of these findings. How well these observations translate to single cell level miRNA expression and their association with cell states remains to be confirmed.

We highlighted an inverse relationship between RNA expression of NPC/OPC genes and MES genes. This observation was also evident in bulk miRNA expression data when comparing NPC/OPC and MES scores and suggests miRNA expression is highly responsive to cell states and likely incorporated into the same gene networks. Although we have preliminary 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 include multiple tumors 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.

In conclusion, our study is the first to implicate the Dlk1-Dio3 and miR-224/452 miRNA clusters as potential regulators of glioblastoma intra-tumoral heterogeneity and may serve as valuable biomarkers for cell state identification.

Data availability

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 (Darmianis dataset), GSE131928 (Nefel and Nefel_10X datasets), GSE102130 (Filbin dataset), and GSE57872 (Patel dataset)^{5,16–18}. Pre-mapped read counts from bulk miRNA and RNA sequencing of glioblastomas were obtained from the GDC Data Portal under project ID CPTAC-3.

Received: 30 October 2023; Accepted: 3 April 2024

Published online: 13 April 2024

References

1. Friedmann-Morvinski, D. Glioblastoma heterogeneity and cancer cell plasticity. *Crit. Rev. Oncog.* **19**, 327–336. <https://doi.org/10.1615/critrevoncog.2014011777> (2014).
2. Parker, N. R., Khong, P., Parkinson, J. F., Howell, V. M. & Wheeler, H. R. Molecular heterogeneity in glioblastoma: potential clinical implications. *Front. Oncol.* **5**, 55. <https://doi.org/10.3389/fonc.2015.00055> (2015).
3. Verhaak, R. G. 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. <https://doi.org/10.1016/j.ccr.2009.12.020> (2010).
4. Wang, Q. et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell* **32**, 42–56. <https://doi.org/10.1016/j.ccell.2017.06.003> (2017).

5. Nefel, C. *et al.* An Integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**, 835–849. <https://doi.org/10.1016/j.cell.2019.06.024> (2019).
6. Smirnova, L. *et al.* Regulation of miRNA expression during neural cell specification. *Eur. J. Neurosci.* **21**, 1469–1477. <https://doi.org/10.1111/j.1460-9568.2005.03978.x> (2005).
7. Ciafre, S. A. *et al.* Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochem. Biophys. Res. Commun.* **334**, 1351–1358. <https://doi.org/10.1016/j.bbrc.2005.07.030> (2005).
8. 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. Bioinform. Syst. Biol.* **2013**, 2. <https://doi.org/10.1186/1687-4153-2013-2> (2013).
9. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12. <https://doi.org/10.14806/ej.17.1.200> (2011).
10. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25. <https://doi.org/10.1186/gb-2009-10-3-r25> (2009).
11. Smith, T., Heger, A. & Sudbery, I. UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499. <https://doi.org/10.1101/gr.209601.116> (2017).
12. Pantano, L., Estivill, X. & Marti, 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. <https://doi.org/10.1093/nar/gkp1127> (2010).
13. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902. <https://doi.org/10.1016/j.cell.2019.05.031> (2019).
14. Faridani, O. R. *et al.* Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.* **34**, 1264–1266. <https://doi.org/10.1038/nbt.3701> (2016).
15. Benetatos, L. *et al.* The microRNAs within the DLK1-DIO3 genomic region: Involvement in disease pathogenesis. *Cell. Mol. Life Sci.* **70**, 795–814. <https://doi.org/10.1007/s00018-012-1080-8> (2013).
16. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401. <https://doi.org/10.1126/science.1254257> (2014).
17. 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. <https://doi.org/10.1016/j.celrep.2017.10.030> (2017).
18. Filbin, M. G. *et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* **360**, 331–335. <https://doi.org/10.1126/science.aao4750> (2018).
19. Crisman, T. J. *et al.* Identification of an efficient gene expression panel for glioblastoma classification. *PLoS One* **11**, e0164649. <https://doi.org/10.1371/journal.pone.0164649> (2016).
20. Fu, M. *et al.* Gene clusters based on OLIG2 and CD276 could distinguish molecular profiling in glioblastoma. *J. Transl. Med.* **19**, 404. <https://doi.org/10.1186/s12967-021-03083-y> (2021).
21. Xiao, Z. *et al.* Holo-Seq: Single-cell sequencing of holo-transcriptome. *Genome Biol.* **19**, 163. <https://doi.org/10.1186/s13059-018-1553-7> (2018).
22. Marziali, G. *et al.* A three-microRNA signature identifies two subtypes of glioblastoma patients with different clinical outcomes. *Mol. Oncol.* **11**, 1115–1129. <https://doi.org/10.1002/1878-0261.12047> (2017).

Author contributions

C.M.S. conducted the analysis wrote the draft manuscript and generated the figures. D.C. edited and reviewed the manuscript. G.H. conceptualised the research and edited and reviewed the manuscript. C.M.S. was a joint PhD student to D.C. and G.H.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-58870-6>.

Correspondence and requests for materials should be addressed to G.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© Crown 2024