NUS
National University
of Singapore

LSM3241 CA1 Report

CHER WEI YUAN (A0155266B)
CHUA YUEN SIONG (A0154961Y)

Abstract

Metastasis is the leading cause of death from cancer. miR-203 is a known micro-RNA that attenuates the migration capabilities of cells by inhibiting the epithelial-mesenchymal transition (EMT). Yet, the transcriptomic response to miR-203 relevant to EMT was not well elucidated. Here, we investigated two Geoquery series (GSE50697 and GSE45121) microarray datasets consisting of treatment samples overexpressing miR-203 and control samples. The objective of microarray analysis was to identify the differentially expressed genes due to miR-203 expression. The tools used to manipulate and statistically evaluate the datasets were RStudio and relevant R packages (GEOquery, affy, limma, hgu133plus2.db and mouse4302.db). Genes identified in RStudio were further evaluated for biological relevance in cancer, EMT and their possible use as cancer treatments using DAVID Bioinformatics and literature review. Subsequently the genes were ranked in accordance to our recommendation for their qPCR validation. In descending order, we recommend the following eight genes: (1) interleukin 6 (IL-6), (2) nidogen 1, (3) LAPTM5, (4) SH3 and SYLF domain containing 1 (SH3YL1), (5) SERPINB3, (6) cyclin D2, (7) chondroitin sulfate N-acetylgalactosaminyltransferase 1 (CSGALNACT1) and (8) parathyroid hormone like hormone (PTHLH). We hope that our recommendations can provide directions to further studies in miR-203, EMT or cancer treatments.

Introduction

MicroRNAs (miRNA) are small RNAs involved in post-transcriptional silencing via specific base pairing with the 3'-untranslated regions (UTR) of mRNA targets (Bartel, 2013). The interaction between miRNAs and mRNA may lead to the cleavage of the latter by Argonaute proteins (Peters & Meister, 2007) or the formation of a hairpin structure at the mRNA 3'-UTR that repress translation (Bartel, 2013). One such miRNA is miR-203. miR-203 is involved in the repression of epithelial-mesenchymal transitions, or EMT (Saini *et al.*, 2011; Viticchiè *et al.*, 2011).

During EMT, the epithelial cells undergo genetic changes and adopt a mesenchymal phenotype (Kalluri & Weinberg, 2010). Upon the completion of EMT, the epithelial cells lose their attachment to the basement membrane and become capable of migrating (Moes *et al.*, 2012). Although EMTs are required for embryonic development and tissue regeneration (Castilla *et al.*, 2011), the post-EMT cells acquire migratory and apoptotic resistant phenotypes which were known to facilitate the development of metastatic cancer (Kalluri & Weinberg, 2010).

Given the importance of EMTs in cancer progression, miR-203, a miRNA inhibiting EMTs, was intensely studied (Peters & Meister, 2007; Lena *et al.*, 2008; Wellner *et al.*, 2009; Castilla *et al.*, 2011; Chiang *et al.*, 2011; Saini *et al.*, 2011; Viticchiè *et al.*, 2011; Zhang *et al.*, 2011; Moes *et al.*, 2012). Observations of repressed miR-203 expression were evident in multiple cancers, including that of the central nervous system, liver, esophagus and blood (T-cell lymphomas and B-cell type acute lymphoblastic leukemia) (Viticchiè *et al.*, 2011). Subsequent studies linked miR-203 activity to the regulation of master transcription factors (SNAI1, ZEB1 and RUNX2) that are involved in the molecular pathways leading to EMTs (Wellner *et al.*, 2009; Saini *et al.*, 2011; Viticchiè *et al.*, 2011).

However, gaps remain in the knowledge of transcriptomic alterations relating to EMTs that were induced by miR-203 repression. Here, we examined two series (GSE50697 and GSE45121) of microarray sample data deposited in GEO (NCBI, 2013) using an assorted set of R packages in RStudio (Core RStudio Team, 2018; R Core Team, 2018). The series contains six microarray samples of SUM159 breast cancer cell lines, half of which were overexpressing miR-203 via retroviral transduction and the remaining were controls with basal miR-203 expression **(**NCBI, 2013).

We aim to statistically identify differentially expressed genes in the miR-203 expressing samples via the empirical Bayes method. Subsequently, biologically relevant genes suspected to be related to EMTs were handpicked through examination using DAVID Bioinformatics Resources 6.8 (Huang, Sherman, & Richard, 2009) and relevant literature review. We hope that our work would help find candidate genes for downstream analysis, such as qPCR.

Materials & Methods

The identification of differentially expressed genes of the series GSE50697 from the NCBI Gene Expression Omnibus database (Edgar *et al.*, 2002) was done in RStudio (version 1.1453; R version 3.5.2). The codes used can be reproduced using the annotated R script uploaded in GitHub (https://github.com/CherWeiYuan/LSM3241-Microarray). Briefly, unprocessed microarray sample data from series GSE50697 were downloaded from the command line with the R package GEOquery **(**Sean & Meltzer, 2007). The downloaded raw data were derived on one platform ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, GPL570) and six samples. A plasmid pBabe puro with a miR-203 insert was injected in three samples of SUM159 breast cancer cell lines to overexpress miR-203. Another three samples of the same cell line were injected with pBabe puro without the miR-203 insert as control.

Conventionally, the microarray samples were represented in the rows of the metadata table (also known as the phenoData). However, in the assayData where the intensities of each probe were recorded, the column represented the microarray samples. Hence, there is a need to ensure that, during the reading of assayData into RStudio the correct metadata corresponded with the assayData. Thus, we first obtain the processed GSE50697 dataset to acquire the treatment conditions from the phenoData. Next, we appended the treatment conditions to the processed GSE50697 dataset and converted the product into a dataframe. Following which, the dataframe was read along with the unprocessed data into R using the affy package (Gautier *et al.*, 2004).

Next, statistical analysis was conducted to identify differentially expressed genes. Using the robust multi-array average (RMA) from the package limma (Wu *et al.*, 2015), the microarray data were background corrected and quantile normalized. The former aims to adjust the fluorescent intensity that was not derived from a specific probe; such noise comes from non-specific binding or adjacent probes within or between array chips (Silver, 2009). By assuming an

exponential distribution for the true signal intensity and a normal distribution for the background noise, the latter can be determined and removed (Silver, 2009). Quantile normalization then takes place (with the assumption that the distribution of the intensities was the same) to allow a fair comparison of intensities between the genes. As multiple probes were used for each gene, the RMA function also summarizes the many intensities for each gene into a single value; the collapsing of multiple values into one is done through median polish (Wu *et al.*, 2015). The value for each gene represents the log-transformed Perfect Match (PM) intensity and can be modelled by:

$$Y_i = T_i \times I(treatment) + N_i \times I(untreated) + e_i$$

where Y_i is the log transformed Perfect Match (PM) intensity for gene i,
T_i is the mean expression level of the treated sample,
N_i is the mean expression level of the untreated sample,
e_i is the remaining background noise and
I is the indicator term (1 for treatment and 0 for control sample).

Hence, again with the limma package, we specified a model matrix and a contrast matrix (treatment minus control). Next, we fitted both matrices to the data. As only a few replicate probes were used to measure the intensity of expression of each gene, the variance for each gene was unreliable (Smyth, 2005). Hence, this variance is adjusted via the empirical Bayes method in the limma package (Wu *et al.*, 2015) using information from other genes (Ritchie *et al.*, 2015). As a result, a more reliable test statistic can be obtained when comparing the mean gene expression levels between the treatment and the control (Smyth, 2005).

With the linear model, we identified genes of interest that were either being upregulated or downregulated by a log fold change value (LFC) of 1.8 and above with a statistical significance level of $p < 0.05$. We then annotated the genes with the R package from Bioconductor (Gentleman, 2004), affymetrix human genome U133 plus 2.0 (Carlson, 2016). Volcano plots were then made with the limma package (Wu *et al.*, 2015). Using the built-in stats package (R Core Team, 2018) with the RColour Brewer package (Neuwirth, 2014), a heatmap, of the identified up- or downregulated genes, was constructed for visualization. For our downstream analysis of the identified genes, we used functional annotation tools on DAVID Bioinformatics Resources 6.8 (Huang, 2008**)** to check if the genes were associated with EMTs.

Following which, we ranked the genes in descending order of importance for further downstream analysis such as qPCR based on three main factors: 1) LFC, 2) biological relevance to EMT and cancer and 3) whether the gene can be used as a possible treatment for cancer. Statistical significance was excluded from the factors for consideration because all genes were identified via a p-value of less than 0.05 criterion.

In addition to the study of GSE50697, we acquired another Geoquery series, GSE45121 (NCBI, 2013) to detect differential gene expression between mouse skin cells with and without miR-203 induction. The analysis conducted for GSE45121 was the same as that for GSE50697 except for the use of the Affymetrix Mouse Genome 430 2.0 Array annotation data to annotate genes

(Carlson, 2016). The R script for GSE45121 analysis was also uploaded in the same Github folder as GSE50697.

Results & Discussion

Identification of differentially expressed genes in GSE50697 statistically

Among all the genes that were obtained from the microarray data, there were a total of 16 genes that had a log fold change (LFC) of more than 1.8 (as a change in gene expression shows a LFC of above 1) (Fig 1). We decided on a lower threshold of 1.8 for LFC to net more genes in order to minimise the occurrence of false negatives (in other words, the biologically relevant genes with lower fold change). We think that this is justified because the genes were still selected through stringent adherence to a low p-value of 0.05. Thus, the differentially expressed genes were still statistically significant. On the other hand, an LFC lower than 1.8 (such as 1.5) would yield 34 genes; such a large number would easily exceed the short list of recommended genes we aimed to provide for expensive downstream analysis (e.g. qPCR). Thus, we focused our attention on the 16 genes with the highest value of LFC.

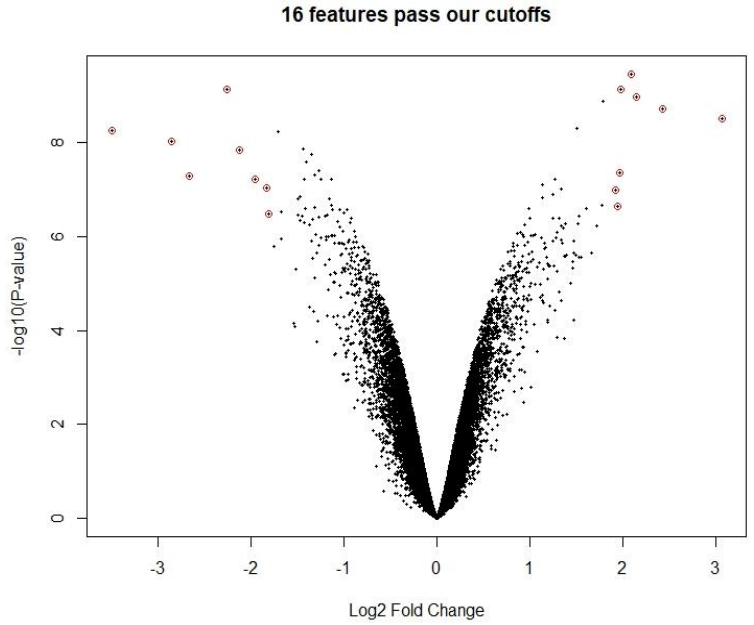

Fig 1. Volcano plot of microarray data with the indication of genes that were above fold change cutoff

Table 1. List of genes above LogFC threshold of 1.8 and their probe IDs

Gene Name	Probe ID	LogFC
SERPINB3	209719_x_at	2.096174
SERPINB3/SERPINB4	210413_x_at	1.984099
LAPTM5	201721_s_at	-2.254152
SERPINB4	211906_s_at	2.143763
SERPINB3	209720_s_at	2.432950

PTHLH	211756_at	3.067195
NID1	202007_at	-3.494914
INTU	228946_at	-2.850568
Cyclin D2	200953_s_at	-2.114548
IL-6	205207_at	1.972971
STK26	218499_at	-2.660353
SH3YL1	204019_s_at	-1.947556
NA	222288_at	-1.827582
NA	1556773_at	1.928811
CSGALNACT1	219049_at	1.948935
NA	1557883_a_at	-1.806322

Based on Table 1, among these 16 selected genes, three probes were not mapped to EntrezID, suggesting that they were not found in the human genome (Maglott, 2010). As we are investigating EMT-associated cancers in humans, these three genes were irrelevant and hence dropped from the downstream analysis. Among the remaining identified genes, three were actually from the same entity: serpin family member 3 (SERPINB3). In addition, serpin family member 4 (SERPINB4) was also detected by two probes. In addition, the serpin 4 duplicate was detected on the same probe, 210413_x_at, as one of the other SERPINB3's duplicates. Through the removal of duplicates, we reduced our list of 16 genes to 11 genes for further examination.

Biological relevance of the differentially expressed genes in GSE50697

According to the Genetic Association Database (GAD) Disease database of DAVID Bioinformatics Resource (Huang, 2008), 8 of the 11 genes were linked to cancer: cyclin D2, interleukin 6 (IL-6), nidogen 1, SERPINB3, parathyroid hormone like hormone (PTHLH), SH3 and SYLF domain containing 1 (SH3YL1), chondroitin sulfate N-acetylgalactosaminyltransferase 1 (CSGALNACT1) and LAPTM5. In descending order of importance based on LFC, biological relevance to EMT and cancer and its potential as a cancer-treatment target, we have ranked the genes we recommend for qPCR validation:

1. IL-6 levels, identified as one of the initiators of cancer cell formation (Fisher *et al.*, 2014) were observed to increase (Fig 2) after treatment with miR-203 (LFC of 1.97). A homolog of IL-6, IL5 prevents apoptosis and was found in high levels in cancer tumours; furthermore, IL5 was known to cause EMT (Gao *et al.*, 2016). Hence, it is possible for IL-6 plays similar support roles for cancer formation. Despite the role of miR-203 in inhibition of EMT and hence cancer (Kalluri & Weinberg, 2010), it is interesting to see miR-203 expression leading to the expression of cancer-causing factors. This conflict was resolved in 2015 when Fisher *et al.* discovered that IL-6 acts as a tumour suppressor by activating T cells which attenuates tumour growth. These facts raise a few questions: does IL-6 cooperate or oppose the cancer-preventing properties of miR-203?

If the latter is true, would reduction of IL-6 further enhance the anti-cancer capability of miR-203? As such, we would highly recommend the confirmation of increased expression of IL-6 via qPCR that may warrant further studies on IL-6.

2. Nidogen 1 induces a signalling pathway leading to EMT (Zhou *et al.*, 2017**)** and hence cancer (Pedrola *et al.*, 2015). In our microarray analysis, it was revealed that the expression of nidogen 1 was lowered after treatment with miR-203. The inhibition of nidogen 1 can prevent EMT and may even drive the reversal of a mesenchymal to an epithelial phenotype (Zhou *et al.*, 2017**)**. This property makes nidogen 1 a possible treatment for cancers. Along with its especially high LFC of -3.49, which suggest the signal may not be artefactual, we highly recommend the qPCR of nidogen 1 for the validation of expression levels.

3. LAPTM5 was found to be downregulated upon treatment with miR-203 in bladder cancer cells (Chen *et al.*, 2017). This downregulation would then prevent cells from entering cell-cycle and at the same time inhibit EMT (Chen *et al.*, 2017). Such a decreased expression level was also observed in our data analysis (LFC of -2.25). Thus, we find LAPTM5 highly relevant for qPCR to confirm that miR-203-induced reduction of LAPTM5 can be generalized to not just bladder cells, but also breast cancer cells.

4. SH3YL1 acts as the receptors for Dock4 ligand to bind to trigger a pathway that will lead to cell migration (Kobayashi *et al.*, 2014). The activation of this pathway is prevalent for breast cancer (Kobayashi *et al.*, 2014). In accordance to the anti-cancer effects of miR-203 through EMT prevention, we observed that the miR-203 treatment reduced the expression of SH3YL1. Hence, there is reason to believe that SH3YL1 plays a role in EMT in response to miR-203, which explains our recommendation for its qPCR analysis.

5. SERPINB3 overexpression was suspected to cause cell migration because it leads to the repression of genes such as E-cadherin, which were required for cells to adhere to one another (Crawford *et al.*, 2014). Furthermore, it would also cause the cell to be able to proliferate without growth factor stimulation, allowing the speed of tumour growth to be increased further (Crawford *et al.*, 2014**)**. Contradictory to the anti-cancer properties of miR-203, all three probes for SERPINB3 showed increased expression levels after miR-203 treatment (LFC of 2.43, Fig 2). However, in another study, SERPINB3 downregulation was found in cancer cells of lung and breast cancer patients (Chou *et al.*, 2012). Given the use of lung cells in that study, we suspect cell-type specific effects of SERPINB3 in both breast and lung cancer cells. Given the potential of SERPINB3 in EMT-induction through reduced cell-adhesion, it would be interesting to address the conflicting data between our results and those from Chou *et al.* (2012).

6. Cyclin D2 plays a role in cell proliferation (Zhao *et al.*, 2015). The elevated expression of cyclin D2 increases the rate at which cells divide (Ladam *et al.*, 2013). In our analysis, samples treated with miR-203 had their levels of cyclin D2 expression downregulated

(LFC of -2.11). When cyclin D2 levels decrease, the cell-cycle halts (Zhao *et al.*, 2015). Although lower levels of cyclin D2 can stop the cell-cycle, it also enables cell migration, which may facilitate EMT (Ladam *et al.*, 2013). Since both up- and down-regulation of cyclin D2 has known potential to cause cancer, where alteration of expression levels does not result in clear outcomes, we do not rank it as highly as the other genes that lack this double-edged nature.

7. CSGALNACT1 is involved in the production of chondroitin sulfate, which is subsequently required by cancer cells for multiple cellular processes such as motility (thus possible involvement in EMTs) and cell-cycle activation (Clausen *et al.*, 2016). However, we observed the upregulation of CSGALNACT1 when miR-203 was overexpressed (LFC of 1.95, Fig 2). There are two possibilities to explain the phenomenon: 1) CSGALNACT1 may have an additional and unknown cancer-attenuating effect, suggesting that it acts through more than just chondroitin sulfate. 2) CSGALNACT1 enhances cancer but its effectiveness was negated by factors caused by miR-203 expression. While the unclear picture warrants further studies, we believe other candidate genes listed here are more relevant in terms of their established (or strong suspected) links to EMT.

8. PTHLH is important to cancer cells as it would cause bones to break down when it is highly expressed (Taipaleenmäki *et al.*, 2015). The nutrients from the destroyed bones would then be reabsorbed and used to fuel the growth of the tumour and lead to eventual migration (Suva *et al.*, 1987). From Fig 2, it was observed that the expression of PTHLH was increased after the cells were treated with miR-203. This outcome was not expected as the upregulation of PTHLH will not only prevent cell death but also make the cancer cells more invasive and undergo EMT (Shen *et al.*, 2004), which is contradictory to the known anti-cancer effects of miR-203. In contrast to the result from our analysis, another study found miR-203 treatment reducing PTHLH expression through the inhibition of RUNX2 pathway (Taipaleenmäki *et al.*, 2015). Thus, the literature conflicts the microarray result we had, and suggests an error in the microarray data whereby PTHLH expression is decreased, rather than increased. This conclusion of a possible error is supported by the fact that both the studies of ours and the literature were done on breast cancer cell lines (SUM159 and MCF-7) (Shen *et al.*, 2004), making cell-type specific effects unlikely. However, it is possible that the cell lines differ in unknown ways that would matter to the outcome of PTHLH expression levels and its contribution to cancer. Overall, we would not recommend qPCR for PTHLH since existing knowledge suggests that our results may be artefactual.

The three remaining genes that showed a significant change in gene expression but were either not mapped to cancer, EMT or both on the GAD Disease database (Huang, 2008). Hence, they were not included in our list of recommended genes. They are SERPINB4, serine/threonine kinase 26 (STK26) and inturned planar cell polarity protein (INTU). Briefly, SERPINB4 had been shown to affect cancer but the effectiveness was limited by the fact that it did not prevent the immune system from attacking the cancer cells and it was not shown to have any effect on EMT

(citation?). Thus, SERPINB4 was not mapped to cancer unlike SERPINB3 which was involved greatly in the formation of cancer and EMT (Izuhara *et al.*, 2018).

STK26 expression levels were successfully lowered upon introduction of miR-203, as seen in Fig 2. STK26 exhibit control over autophagy, yet when excessive autophagy was induced, tumour formation may occur (Huang *et al.*, 2017). As STK26 plays no known role in EMT, it is irrelevant to our recommendation of miR-203-induced EMT genes. Another gene, INTU is required for cells to form cilia and tend to be overexpressed in cancer cells (Eguether & Hahne, 2018). Similarly, the treatment of miR-203 caused the expression of the gene to be lowered but the gene itself did not have any direct impact on tumour formation or EMT. Thus, INTU was excluded from our recommendations.

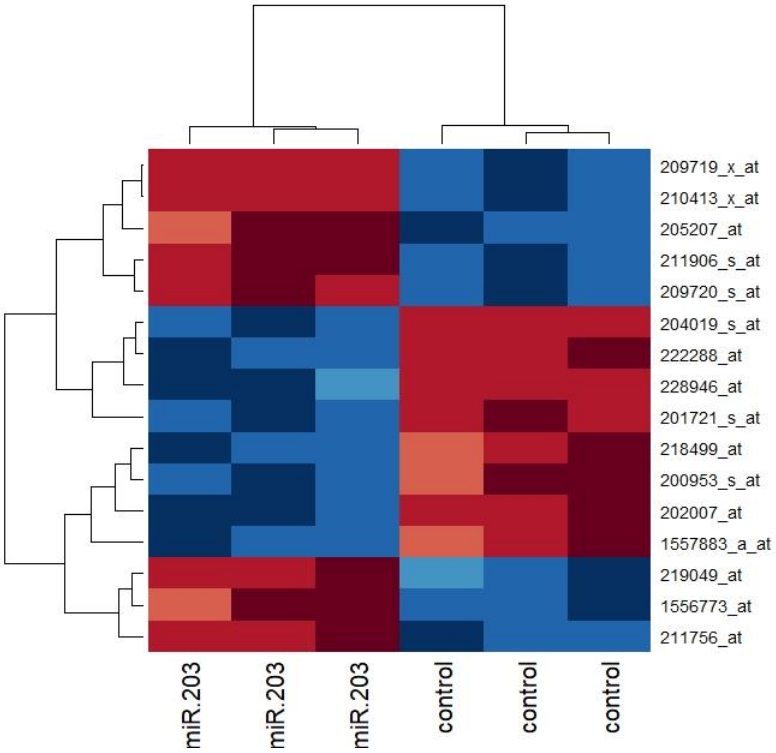

Fig 2. Heatmap of genes that were above the cutoff

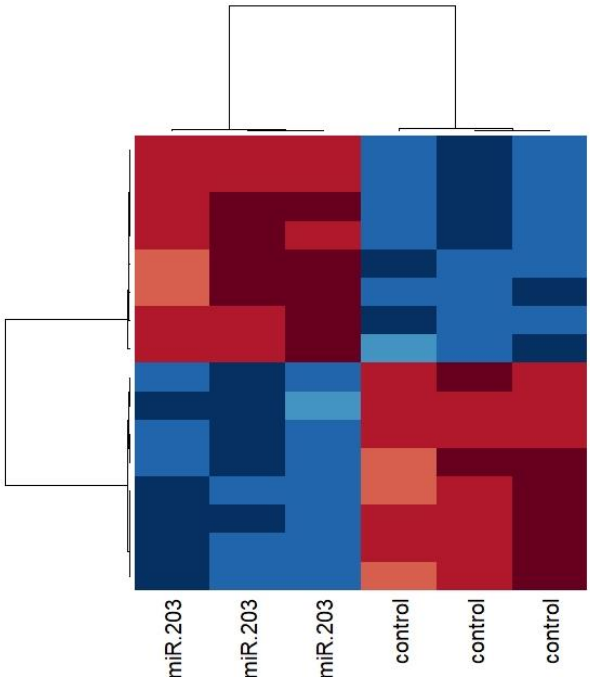

Fig 3. Heatmap after reclustering

In addition to the analysis done on GSE50697, we acquired GSE45121 (NCBI, 2013) to detect differential gene expression between mouse skin cells with and without miR-203 induction. Although the microarray dataset was not derived from human cells, any differentially expressed genes between the treatment and control could have a homolog in humans that was worth further investigation. We intended to obtain the sequences for the probes of differentially expressed genes and conduct a NCBI BLASTN to find human homologs. However, among the top ten genes ranked by the B-statistic in the eBayes model, none of the genes had a LFC higher than 1.8 (the highest was 1.44 for Rho GTPase binding protein 2) and an adjusted p-value lower than 0.05. Given the low statistical significance in addition to the non-human source of data, we cannot justify the recommendation of any genes from GSE50697.

We identified one limitation with our work. We noticed that better background adjustment methods could be formulated as currently, background adjustment only correctly identifies slightly more true positives while the false positive numbers were not reduced (Qin *et al.*, 2013). These could affect the results greatly as other genes that may show statistically significant differences that were greater than the genes that were identified, resulting in the identification of artefactual genes. However, such a limitation may be counteracted by our attempt to find biological relevance to the recommended list of genes.

Conclusion

We recommended 8 genes for downstream qPCR analysis to validate the microarray differential gene expression results. Our work is important because the qPCR of all eight genes (or the rest

of the genes differentially expressed) may be impossible due to budget restrictions. Following qPCR, the confirmation of the microarray results can then allow the further study of the effects of the genes on miR-203-induced EMT. Such validations are important because, as shown in our discussion, the microarray results may not agree with the current literature. These conflicts may be due to errors in the microarray or its analysis but may reflect true biological relevance, such as cell-type specific effects. Hence, we hope our restricted list of genes may prove relevant in future research on EMT and miR-203.

References

Bartel, D. (2013). MicroRNA Target Recognition and Regulatory Functions. *Cell*, *136*(2), 215–233. https://doi.org/10.1016/j.cell.2009.01.002.MicroRNA

Carlson, M. (2016). Affymetrix Mouse Genome 430 2.0 Array annotation data (chip mouse4302). Retrieved from http://bioconductor.org/packages/release/data/annotation/html/mouse4302.db.html

Carlson, M. (2016). hgu133plus2.db: Affymetrix Human Genome U133 Plus 2.0 Array annotation data (chip hgu133plus2).

Castilla, M., Moreno-bueno, G., Romero-p, L., Vijver, K. Van De, Biscuola, M., Cano, A., & Oliva, E. (2011). Micro-RNA signature of the epithelial – mesenchymal transition in endometrial carcinosarcoma. *Journal of Pathology*, (October 2010), 72–80.

Chiang, Y., Song, Y., Wang, Z., Chen, Y., Yue, Z., Xu, H., … Liu, Z. (2011). Aberrant Expression of miR-203 and Its Clinical Significance in Gastric and Colorectal Cancers. *Journal of Gastrointestinal Surgery*, *15*(1), 63–70. https://doi.org/10.1007/s11605-010-1367-8

Chou, R.-H., Chou, R.-H., Wen, H.-C., Wen, H.-C., Liang, W.-G., Liang, W.-G., … Chang, W.-S. W. (2011). Suppression of the invasion and migration of cancer cells by SERPINB family genes and their derived peptides. *Oncology Reports*, *27*(1), 238–245. https://doi.org/10.3892/or.2011.1497

Clausen, T. M., Pereira, M. A., Al Nakouzi, N., Oo, H. Z., Agerbaek, M., Lee, S., … Salanti, A. (2016). Oncofetal Chondroitin Sulfate Glycosaminoglycans Are Key Players in Integrin Signaling and Tumor Cell Motility. https://doi.org/10.1158/1541-7786.MCR-16-0103

Core RStudio Team. (2018). RStudio: Open source and enterprise-ready professional software for R. Boston. Retrieved from https://www.rstudio.com/

Crawford, H. C., Ullman, E., Wu, S., Chen, E. I., Zhang, J., Sheshadri, N., … Pan, J.-A. (2014). SCCA1/SERPINB3 Promotes Oncogenesis and Epithelial-Mesenchymal Transition via the

Unfolded Protein Response and IL6 Signaling. *Cancer Research*. https://doi.org/10.1158/0008-5472.can-14-0798

Edgar, R., Domrachev, M., & Lash, A. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, *30*(1), 207–210. https://doi.org/10.1093/nar/30.1.207

Eguether, T., & Hahne, M. (2018). Mixed signals from the cell's antennae: primary cilia in cancer. *EMBO Reports*, *19*(11), e46589. https://doi.org/10.15252/embr.201846589

Xiaoli, S., Lihui, Q., & Miriam, F. (2004). PTH-related protein enhances MCF-7 breast cancer cell adhesion, migration, and invasion via an intracrine pathway. *Experimental Cell Research*. https://doi.org/DOI: 10.1016/j.yexcr.2003.11.028

Gao, J., Zhao, S., & Halstensen, T. S. (2016). Increased interleukin-6 expression is associated with poor prognosis and acquired cisplatin resistance in head and neck squamous cell carcinoma. *Oncology Reports*, *35*(6), 3265–3274. https://doi.org/10.3892/or.2016.4765

Gautier, L., Cope, L., Bolstad, B. M., & Irizarry, R. A. (2004). Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, *20*(3), 307–315. https://doi.org/10.1093/bioinformatics/btg405

Gentleman, R. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, *5*(10), R80. https://doi.org/10.1186/gb-2004-5-10-r80

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, *37*(1), 1–13. https://doi.org/10.1093/nar/gkn923

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Huang, T., Kim, C. K., Alvarez, A. A., Pangeni, R. P., Wan, X., Song, X., … Cheng, S.-Y. (2017). MST4 Phosphorylation of ATG4B Regulates Autophagic Activity, Tumorigenicity, and Radioresistance in Glioblastoma. *Cancer Cell*, *32*(6), 840–855.e8. https://doi.org/10.1016/j.ccell.2017.11.005

Izuhara, K., Yamaguchi, Y., Ohta, S., Nunomura, S., Nanri, Y., Azuma, Y., … Aihara, M. (2018). Squamous Cell Carcinoma Antigen 2 (SCCA2, SERPINB4): An Emerging Biomarker for Skin Inflammatory Diseases. *International Journal of Molecular Sciences*, *19*(4), 1102. https://doi.org/10.3390/ijms19041102

Kalluri, R., & Weinberg, R. (2010). The basics of epithelial-mesenchymal transition. *The Journal of Clinical Investigation*. https://doi.org/10.1172/JCI39104.1420

Kobayashi, M., Harada, K., Negishi, M., & Katoh, H. (2014). Dock4 forms a complex with SH3YL1 and regulates cancer cell migration. *Cellular Signalling*, *26*(5), 1082–1088. https://doi.org/10.1016/J.CELLSIG.2014.01.027

Ladam, F., Damour, I., Dumont, P., Kherrouche, Z., de Launoit, Y., Tulasne, D., & Chotteau-Lelievre, A. (2013). Loss of a Negative Feedback Loop Involving Pea3 and Cyclin D2 Is Required for Pea3-Induced Migration in Transformed Mammary Epithelial Cells. *Molecular Cancer Research*, *11*(11), 1412–1424. https://doi.org/10.1158/1541-7786.MCR-13-0229

Lena, A. M., Shalom-Feuerstein, R., di Val Cervo, P. R., Aberdam, D., Knight, R. A., Melino, G., & Candi, E. (2008). miR-203 represses "stemness" by repressing ΔNp63. *Cell Death and Differentiation*, *15*(7), 1187–1195. https://doi.org/10.1038/cdd.2008.69

Moes, M., Bechec, A. Le, Crespo, I., Laurini, C., Halavatyi, A., Vetter, G., … Friederich, E. (2012). A Novel Network Integrating a miRNA-203 / SNAI1 Feedback Loop which Regulates Epithelial to Mesenchymal Transition. *PLoS ONE*, *7*(4). https://doi.org/10.1371/journal.pone.0035440

Munkley, J., Vodak, D., Livermore, K. E., James, K., Wilson, B. T., Knight, B., … Elliott, D. J. (2016). Glycosylation is an Androgen-Regulated Process Essential for Prostate Cancer Cell Viability. *EBioMedicine*, *8*, 103–116. https://doi.org/10.1016/j.ebiom.2016.04.018

NCBI. (2013). GEO Series GSE45121. Retrieved from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45121

NCBI. (2013). GEO Series GSE50697. Retrieved from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50697

Neuwirth, E. (2005). The RColorBrewer Package. *Cran.R-Project.Org*, 3–6. Retrieved from http://ftp.auckland.ac.nz/software/CRAN/doc/packages/RColorBrewer.pdf

Nuylan, M., Kawano, T., Inazawa, J., & Inoue, J. (2016). Down-regulation of LAPTM5 in human cancer cells. *Oncotarget*, *7*(19), 28320–8. https://doi.org/10.18632/oncotarget.8614

Peters, L., & Meister, G. (2007). Argonaute Proteins: Mediators of RNA Silencing. *Molecular Cell*, *26*(5), 611–623. https://doi.org/10.1016/j.molcel.2007.05.001

Qin, L. X., Tuschl, T., & Singer, S. (2013). An empirical evaluation of normalization methods for microRNA arrays in a liposarcoma study. *Cancer Informatics*. https://doi.org/10.4137/CIN.S11384

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://cran.r-project.org/bin/windows/base/

Saini, S., Majid, S., Yamamura, S., Tabatabai, L., & Suh, S. O. (2011). Regulatory Role of mir-203 in Prostate Cancer Progression and Metastasis. *Clinical Cancer Research*, (7), 5287–5299. https://doi.org/10.1158/1078-0432.CCR-10-2619

Sean, D., & Meltzer, P. S. (2007). GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, *23*(14), 1846–1847. https://doi.org/10.1093/bioinformatics/btm254

Silver, J. D., Ritchie, M. E., & Smyth, G. K. (2009). Microarray background correction: Maximum likelihood estimation for the normal-exponential convolution. *Biostatistics*, *10*(2), 352–363. https://doi.org/10.1093/biostatistics/kxn042

Smyth, G. K. (2005). Limma: linear models for microarray data BT - Bioinformatics and Computational Biology Solutions Using R and Bioconductor. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, 397–420. https://doi.org/10.1007/0-387-29362-0_23

Suva, A. L. J., Winslow, G. A., Wettenhall, R. E. H., Hammonds, R. G., Moseley, J. M., Rodda, C. P., … Wood, W. I. (2019). A Parathyroid Hormone-Related Protein Implicated in Malignant Hypercalcemia : Cloning and Expression Published by : American Association for the Advancement of Science Stable URL : https://www.jstor.org/stable/1699896 REFERENCES Linked references are avai.

Taipaleenmäki, H., Browne, G., Akech, J., Zustin, J., van Wijnen, A. J., Stein, J. L., … Lian, J. B. (2015). Targeting of Runx2 by miR-135 and miR-203 Impairs Progression of Breast Cancer and Metastatic Bone Disease. *Cancer Research*, *75*(7), 1433–44. https://doi.org/10.1158/0008-5472.CAN-14-1026

Vaes, E., Khan, M., & Mombaerts, P. (2014). Statistical analysis of differential gene expression relative to a fold change threshold on NanoString data of mouse odorant receptor genes. *BMC Bioinformatics*, *15*, 39. https://doi.org/10.1186/1471-2105-15-39

Viticchiè, G., Lena, A. M., Latina, A., Formosa, A., Gregersen, L. H., Lund, A. H., … Melino, G. (2011). MiR-203 controls proliferation , migration and invasive potential of prostate cancer cell lines. *Cell Cycle*, (December 2013). https://doi.org/10.4161/cc.10.7.15180

Wellner, U., Schubert, J., Burk, U. C., Schmalhofer, O., Zhu, F., Sonntag, A., … Keck, T. (2009). The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nature Cell Biology*, *11*(12), 1487–1495. https://doi.org/10.1038/ncb1998

Wu, D., Smyth, G. K., Ritchie, M. E., Law, C. W., Phipson, B., Hu, Y., & Shi, W. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007

Xiao, M., Li, J., Li, W., Wang, Y., Wu, F., Xi, Y., … Yu, W. (2017). MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biology*, *14*(10), 1326–1334. https://doi.org/10.1080/15476286.2015.1112487

Xiaoli, S., Lihui, Q., & Miriam, F. (2004). PTH-related protein enhances MCF-7 breast cancer cell adhesion, migration, and invasion via an intracrine pathway. *Experimental Cell Research*. https://doi.org/DOI: 10.1016/j.yexcr.2003.11.028

Zhang, Z., Zhang, B., Li, W., Fu, L., Fu, L., Zhu, Z., & Dong, J. T. (2011). Epigenetic Silencing of miR-203 Upregulates SNAI2 and Contributes to the Invasiveness of Malignant Breast Cancer Cells. *Genes and Cancer*, *2*(8), 782–791. https://doi.org/10.1177/1947601911429743

Zhao, S., Han, J., Zheng, L., Yang, Z., Zhao, L., & Lv, Y. (2015). MicroRNA-203 Regulates Growth and Metastasis of Breast Cancer. *Cellular Physiology and Biochemistry*, *37*(1), 35–42. https://doi.org/10.1159/000430331