Label Free Proteomic Analysis Reveals Large Dynamic Changes to the Cellular Proteome Upon Expression of the miRNA-23a-27a-24-2 MicroRNA Cluster.

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Biochemistry and Cell Biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>bcb-2019-0014.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>17-Apr-2019</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Piragasam, Ramanaguru; University of Alberta, Biochemistry Hussain, S; University of Alberta, Medicine Chaulk, Steven; University of Alberta, Biochemistry Siddiqi, Zaeem; University of Alberta, Medicine Fahlman, Richard P.; University of Alberta, Biochemistry</td>
</tr>
<tr>
<td>Keyword:</td>
<td>miRNA, miR-23~24, label free proteomics, miR23a, miR27a</td>
</tr>
<tr>
<td>Is the invited manuscript for consideration in a Special Issue? :</td>
<td>Ribowest RNA</td>
</tr>
</tbody>
</table>

https://mc06.manuscriptcentral.com/bcb-pubs
Label Free Proteomic Analysis Reveals Large Dynamic Changes to the Cellular Proteome Upon Expression of the miRNA-23a-27a-24-2 MicroRNA Cluster.

Ramanaguru S. Piragasam, S. Faraz Hussain, Steven G. Chaulk, Zaeem A. Siddiqi and Richard P. Fahlman

Affiliations
R.S. Piragasam, S.G. Chaulk and R.P. Fahlman. Department of Biochemistry, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, AB, Canada
F. Hussain and Z.A. Siddiqi. Department of Medicine, Faculty of Medicine & Dentistry, University of Alberta.
R.P. Fahlman. Department of Oncology, Faculty of Medicine & Dentistry, University of Alberta.
Corresponding author. Richard P. Fahlman (email: rfahlman@ualberta.ca)

Key Words: miRNA, miR23a, miR27a, miR24-2, miR-23~24, label free proteomics.

Abstract
In deciphering the regulatory networks of gene expression controlled by the small non-coding RNAs known as microRNAs (miRNAs), a major challenge has been with the identification of the true mRNA targets by these RNAs within the context of the enormous numbers of predicted targets for each of these small RNAs. To facilitate the system wide identification of miRNA targets a variety of system wide methods, such as proteomics, have been implemented. Here we describe the utilization of quantitative label-free proteomics and bioinformatics to identify the most significant changes to the proteome upon expression of the miR-23a-27a-24-2 miRNA cluster. In light of recent work leading to the hypothesis that only the most pronounced regulatory events by miRNAs may be physiologically relevant, our data reveals that label-free analysis circumvents the limitations of proteomic labeling techniques that limit the maximum differences that can be quantified. The result of our analysis identifies a series of novel candidate targets that are reduced in abundance by more than an order of magnitude upon the expression of the miR-23a-27a-24-2 cluster.
Introduction

For decades since their discovery (Lee et al. 1993), gene expression regulation by microRNAs (miRNAs) has been widely investigated with respect to their mechanisms of gene regulation and their networks of mRNA targets. A major mechanism of gene regulation by these small RNAs is their guiding of the RISC complex to target mRNAs by partial base pairing to target sequences within the mRNA sequence (Fabian and Sonenberg 2012). As a result, to unravel the complexity of miRNA:mRNA regulation networks, numerous investigations have been pursued to identify the targets of individual miRNAs. A summation of many these investigations has been archived in the validated miRNA target database, miRTarBase (Chou et al. 2018). Despite this extensive archive of validated targets, the number of predicted miRNA targets by bioinformatic algorithms, as those utilized by TargetScan (Agarwal et al. 2015), far surpasses the number of validated targets.

While bioinformatic tools have proven invaluable in the identification of potential miRNA targets, these targets still require experimental validation as target prediction algorithms predict many false positive identifications (Pinzon et al. 2017). In addition, we have an incomplete molecular understanding the targeting of mRNAs by miRNAs, for example miR-23a and miR-23b differ by a single nucleotide outside of their common seed region and thus are predicted to have identical miRNA targets by bioinformatic analysis. On the other hand, laboratory experiments have revealed that in some instances these two miRNAs share targets, such as Pdcd4 (Hu et al. 2017) but other examples have revealed unique targets for these miRNAs (Li et al. 2016).

Regarding the biological function of miRNAs, an underappreciated aspect is the occurrence of multiple miRNA in the primary miRNA transcript (pri-miRNA). While not all miRNAs encoded in close proximity in the genome are transcribed as a cluster on a single pri-RNA transcript, many are (Chaulk et al. 2016). The forms of miRNA clusters include those encoded in the introns of coding mRNAs, such as the miR-17~92 cluster (He et al. 2005) and those expressed from their own promoter, such as the miR-23a-27a-24-2 (miR-23~24) cluster (Kong et al. 2010). This linked transcription of the miRNAs within a cluster dictates their coordinated expression. Mechanisms have been identified that can lead to differential miRNA maturation of the individual miRNA from the pri-miRNAs, for example differences in maturation of the miRNA from the miR-17-92 cluster as a result of the folding of the pri-miRNA transcript (Chaulk et al. 2014a; Chaulk et al. 2011). Despite this, the miRNAs of a pri-miRNA transcript must be considered as a biological unit. For the miR-23~24, synergistic functions between miR-23a, miR-27a and miR-24-2 have been identified bioinformatically (Chhabra et al. 2010) and by comprehensive proteomic analysis (Ludwig et al. 2016).

Overshadowing the rapid expansion in the literature regarding various miRNA:mRNA interactions, recent investigations and hypotheses are suggesting that much research on miRNA targeting maybe highly over interpreted (Seitz 2017). Investigations have demonstrated that the often reported small changes in gene expression, often under two fold, by miRNAs is often less than the variability observed in gene expression between individuals (Pinzon et al. 2017). As opposed to the large scale changes in gene expression initially observed with the initial identification of miRNAs (Lee et al. 1993), it is suggested we may need to seriously reconsider the interpretation of small changes in mRNA regulation that is often reported (Seitz 2017).

In light of this issue regarding the importance of identifying the most significant changes in gene expression by miRNAs to identify the potentially most physiologically relevant miRNA:mRNA associations, we are investigating the utility of label free proteomics to quantify
the most significant changes to the proteome upon miRNA expression. While proteomic labeling methods, such as iTRAQ isobaric tags, have been demonstrated to be of high utility to identify changes to the cellular proteome as a result of miRNA expression (Ludwig et al. 2016), the magnitude of changes are typically underestimated as a result of experimental artifacts, that are not widely discussed, that leads to a suppression of the quantified differences observed between samples (Ow et al. 2009; Wang et al. 2012). Here we report on the label free quantification of the proteome changes that result upon the expression of the miR-23~24 cluster, where previous reports using iTRAQ almost exclusively only observed changes of under two-fold.

Materials and Methods

miR-23~24 expression in Cell Culture

To express the miR-23~24 cluster, the genomic sequence including the 2 kb sequence downstream of the cluster was cloned into a pcDNA 3.1(+) vector. HEK293T cells obtained from ATCC were grown in DMEM supplemented with 10% fetal bovine serum. Transfection of the miR-23~24 plasmid or empty vector control were performed using the calcium phosphate-based method as described previously (Jordan et al. 1996) to a monolayer of cells in a 10 cm dish. 48 hours after transfection the cells were lysed for either RNA or protein analysis as described below.

RNA Analysis

Total RNA was isolated using Trizol (Invitrogen) for Northern blot analysis as we have previously described (Chaulk et al. 2014b; Chaulk et al. 2011). For Northern blot analysis the following DNA probes purchased from IDT were used for each miRNA. miR-23a-3P: 5’-ggaaatccctggcaatgtgat-3’. miR-27a-3P: 5’-gcggaacttagccactgtgaa-3’. miR-24-2-3P: 5’-ctgttcctgcatgccacagcc-3’.

Proteomic Analysis

For proteomic analysis, the cells in a 10 cm dish were harvested 48 hours after transfection by lysis in 150 µL denaturing lysis buffer (50 mm Tris, pH 6.8, 8% glycerol (v/v), 1% SDS (w/v), 0.125% β-mercaptoethanol (v/v), 1 mM PMSF, and 1 µg/ml of leupeptin). To facilitate pipetting, the samples were then sonicated to shear the genomic DNA in the samples. 30µL of lysate of each sample was resolved by 10% SDS-PAGE and then visualized with Coomassie Blue staining. Each lane was excised and subsequently cut into 14 equal bands where each band contains proteins from a different molecular weight range. Each gel fraction was subjected to in-gel tryptic digestion as previously described (Khan et al. 2015) and resulting peptides were dried and resuspended in 60µL of 0.2% formic acid in 5% acetonitrile (ACN). Digested peptides were analyzed by LC-MS/MS using a ThermoScientific Easy nLC-1000 in tandem with a Q-Exactive Orbitrap mass spectrometer. Each sample (5µL) was resolved using a 120 minute gradient (0-45% Buffer B; Buffer A : 0.2 formic acid in 5% ACN, Buffer B: 0.2% formic acid in ACN) on a 2 cm Acclaim 100 PepMap Nanoviper C18 trapping column in tandem with a Thermo EASY-Spray column (PepMap® RSLC, C18, 3µm, 100Å, 75µm x 150mm). For data dependent analysis, full scans were acquired at 35000 resolution at a range of 400-200 m/z while 17500 resolution was used for MS/MS scans. Only top 15 ions with +2 and +3 charges were selected for MS/MS with 10-second dynamic exclusion applied to prevent continuous reanalysis of abundant peptides. Following data acquisition, raw data files were compiled for
each gel lane and searched with Proteome Discoverer 1.4’s SEQUEST search algorithm using the reviewed, non-redundant *homo sapiens* complete proteome retrieved from UniProtKB. The search parameters and quantification were as previously described (Kramer et al. 2017).

**Data Analysis**

The false discovery rates (FDRs) for the identifications of the analyzed samples were as follows: “vector control” samples’ actual relaxed FDRs for individual peptides were 0.0438, 0.0416, and 0.0430 while the actual strict FDRs were 0.0068, 0.0067, and 0.0067 for replicates 1-3 respectively; “miR23~24” samples’ actual relaxed FDRs were 0.0427, 0.0358, and 0.0394 while the actual strict FDRs were 0.0070, 0.0078, and 0.0070 for replicates 1-3 respectively. Subsequent analysis was carried out in Microsoft Excel. Extracted Ion Chromatogram (EIC) was used as a measure of protein abundance with only proteins with EIC>0 in ≥1 sample were used for comparative data analysis. Relative Total Ion Current (TIC) was calculated by summing the EIC for all the proteins identified in a sample. Then each proteins’ EIC were divided by TIC to obtain the “proportion of total” value per sample. To determine changes in abundance in two sample sets (miR-23~24 vs. vector control), a two-tailed, heteroscedastic Student’s t-test was applied. Resulting p-values were sorted and uploaded to the “q-value estimation for FDR control” web utility (qvalue.princeton.edu) (Storey and Tibshirani 2003) to generate estimates of the FDR(q-values). The complete set of proteomic data is provided in Supporting Information Table 1. Functional analysis of proteins with p<0.01 were analyzed with Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 6.8 (https://david.ncifcrf.gov/) (Huang da et al. 2009) and enriched for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway identifiers. Predicted miR23~24 targets were obtained from TargetScanHuman version 7.2 (http://www.targetscan.org/vert_72/) (Agarwal et al. 2015) while the validated miR23~24 targets were obtained from miRTarBase version 7.0 (http://mirtarbase.mbc.nctu.edu.tw/php/index.php) (Chou et al. 2018).

**Parallel Reaction Monitoring**

Proteins of interest were quantified via Parallel Reaction Monitoring on a Q-Exactive Orbitrap mass spectrometer coupled with Thermo Scientific Easy nLC-1000. Samples were prepared as described above with 40µL lysate were resolved to approximately 30% of gel length by SDS-PAGE then subjected to in-gel trypsin digestion. The resulting samples were then combined to be analyzed in a single LC-MS/MS run. Multiple peptides per protein (Supp. Table 4) were included for method building with acquisition beginning with a high-resolution full scan (35000) followed by individual precursor isolation by quadrupole (isolation window 1.6 m/z), HCD fragmentation (NCE 27%) and analyzed by the orbitrap analyzed at 17500 resolution. Raw data were imported and analyzed using Skyline software (MacLean et al. 2010). Peptides’ chromatographic peak identity were confirmed by importing Proteome Discoverer result files obtained from the PRM raw data to Skyline and aligning the peptide’s retention time to raw chromatogram.

**Results**

**MicroRNA Cluster Expression**

To investigate the miR-23~24 cluster the entire 2.2 kilo-base sequence was cloned into a pcDNA vector. This plasmid or a vector control were transfected into HEK 293T cells and the total RNA
from these cells were analyzed by Northern Blot analysis for miR23a (Figure 1A). In addition, the total RNA from a Hela cells and a series of breast cancer derived cell lines was also analyzed for endogenous miR23a levels. As seen in Figure 1A, the expression of the miR-23~24 cluster from the plasmid in HEK293T cells lead to miR23a levels that are comparable to endogenous levels of the miRNA observed in MDA-MB-468 and MDA-MB-231 cells. As with the absence of detectable miR23a in HEK293T cells, miR27a and miR24-2 are only detectible in these cells upon transfection of the vector for miR-23~24 expression (Figure 1B).

Proteomic Analysis

To investigate the proteome wide changes in cells upon expression of the miR-23~24 cluster, HEK293T cells were transfected with the miR-23~24 expression vector or an empty vector control. Triplicate samples of the controls and the miR-23~24 cluster expressing cells were then lysed and analyzed by Gel-LC-MS/MS. The analysis identified a total of 4,349 proteins (Supp. Table 1) that were quantified for their relative intensity by quantifying the extracted ion chromatograms (EIC) for the three most intense tryptic peptide observed for each protein and normalizing this data to the total ion current quantified for the sample. The triplicate relative EIC intensities of each protein were compared by individual Student’s t-tests. The complete set of statistical comparisons of the normalized data is listed in Supp. Table 2.

The reproducibility of the label free quantification is relatively high. The comparison of the quantified data from two individual control samples reveals only minimal run-to-run variability for proteins quantified below the relative ion intensity of $1 \times 10^{-3}$ (Figure 2A), while the majority of the data exhibits reproducible quantification. In contrast, when the average relative ion intensities for the three replicate vector control samples are plotted against the average relative ion intensities of the miR-23~24 expressing cells, an overall reduction in the levels of many proteins is observed with the number of proteins observed below the diagonal trendline (Figure 2B).

For a more comprehensive comparison of the miR-23~24 expressing and the vector control cells, including the statistical analysis between the datasets, the fold change of each protein’s relative average ion intensities are depicted as a volcano plot in Figure 2C. Proteins uniquely observed in a single experimental condition were assigned a log$_2$ value of $\pm 10$ to facilitate their visualization on the log scale plot. In this case, proteins only observed once or twice in a particular sample and absent in the other exhibit high p-values while proteins repetitively observed in only a single experimental condition result in a low p-value, for example Orc6 was uniquely observed in all three vector control samples but was not detected in any sample from the miR-23~24 expressing cells.

Overall the analysis again reveals an overrepresentation of down regulated proteins upon miR-23~24 cluster expression, as viewed by the asymmetry of the volcano plot with more proteins being observed to the left of the Y-axis. To query the quality of the data, a series of proteins often utilized as loading controls were analyzed for their relative extracted ion intensities. The ion intensities for Gapdh, ActB, GPI, and CFL1 are shown in Figure 2D and none reveal an observable difference between the control or miR-23~24 cluster expressing cells. For comparison with these invariant protein controls are plotted two histone proteins, Hist1H2BK (Hist H2B type 1-K) and H2AFX (H2A histone family member X), where Hist1H2BK again exhibits no observable variation between samples but nearly a ten-fold decrease is observed for H2AFX. H2AFX was predicted to be down regulated upon expression...
of the miR-23~24 cluster as it is a previously reported target for miR-24-2 (Srivastava et al. 2011).

**Figure 2E** summarizes the observed number of up and down regulated proteins at p-value cut-offs of 0.05, 0.01 and 0.001. As a result of the multiple testing problem when comparing large datasets (Storey and Tibshirani 2003), the q-values for each of these p-value cut-offs were determined as an estimate of the false discovery rate. The corresponding q-values for these p-values were 0.139, 0.099 and 0.094 respectively.

**Functional analysis of the altered proteome**

For a global analysis of the proteome wide changes observed in the cells upon miR-23~24 cluster expression, bioinformatic analysis was performed on the function of the proteins that exhibited a change in abundance. For this analysis, proteins that met the criteria of a p-value of <0.01 were analyzed. The proteins were analyzed with the DAVID functional annotation tool for KEGG pathway identifiers (Huang da et al. 2009) to identify potential enrichments of functions for the proteins observed to change in abundance. As seen in Table 1, a series of functional groups were identified including ribosomes, RNA processing and metabolic processes including glutathione metabolism and oxidative phosphorylation.

**Changes in Ribosomal Proteins**

With ribosomal proteins topping the list from the DAVID analysis, a closer look at all of the ribosomal proteins observed in the data was performed. **Figure 3A** reveal the quantification of ribosomal proteins observed in the control and miR-23~24 cluster expressing cells, where a general global down regulation is observed apart from a few ribosomal proteins such as RpsA and Rpl7L1. This observed general overall reduction in ribosomal proteins, which may reflect an overall reduction in ribosomes, upon miR-23~24 cluster expression is consistent with the global reduction in the large number of proteins observed (Figure 2).

**Validation of Changes in Ribosomal Protein Abundance.**

With our label free shotgun proteomic analysis revealing large fold changes in a number of proteins upon the expression of the miR-23~24 cluster, some of these changes were validated by an alternative method. For this we utilized the quantitative approach of parallel reaction monitoring (Gallien et al. 2012) to quantify the changes in abundance of select proteins. The data in **Figure 3B**, reveals the quantification of Rps8, Rps9 and GAPDH from both the initial three shotgun analysis and from three PRM analysis of control and miR-23~24 expressing cells. Of methodological note, the PRM analysis was performed on whole cellular lysate and not lysates pre-fractionated by SDS-PAGE. As seen in **Figure 3B**, the data between the two analytical methods are in good agreement regarding the magnitudes of change in protein abundance. As predicted, the PRM data resulted in higher precision quantification of these changes.

**Comparison of Proteomic Data with miRNA Predicted Targets**

With the large-scale proteome changes observed upon miR-23~24 cluster expression being a result of both direct targeting by miRNAs and indirect as a result of the miRNA targets regulating the expression of others, a comparison of the proteomic changes observed was made with both the predicted targets and validated targets listed in the Target Scan and miRTarBase bases. For this analysis the proteins that were observed to decrease or increase in abundance in the proteome datasets were determined whether they were targets in either of the databases.
complete comparison is listed in Supp. Table 3, but the summary of the data in Figure 4 strikingly reveals a minimal overlap of the down regulated proteins with the miRNA target datasets. In addition, numerous proteins observed to increase in abundance are predicted to be targets of either miR23a, miR27a or miR24-2.

Top Candidate Targets of the MiR-23~24 Cluster

While many of the down regulated proteins are likely a result of the indirect activity of the miRNAs within the miR-23~24 cluster, the intersection of predicted targets and the down regulated proteins by proteomic analysis was used to identify candidate targets. While the overlapping still leads to a significant number of candidates, 89 with a p-value of <0.05 (Figure 4), we have also filtered the data for the largest magnitude change in abundance. Figure 5 displays the quantified data for 11 proteins that were either only observed in the vector control cells or exhibited at least a 10-fold reduction in expression, while also exhibiting a p-value of less than 0.005.

Discussion

Here we have utilized the over expression of the miR-23~24 cluster in HEK 293T cells to demonstrate the utility of label free proteomics to identify large scale alterations to the cellular proteome. While the expression of genes from a CMV promoter in these cells can lead to artificially high levels of gene expression, in this model it is apparent from the miRNA expression analysis in Figure 1 that expression of the miR-23~24 cluster leads to high levels of mature miRNAs but is comparable to levels observed endogenously in widely utilized breast cancer derived cells lines, such as MDA-MB-231 and MDA-MB-468 cells. These high levels may be relevant in some aspects of breast cancer as the miR-23~24 cluster has been linked to the progression of these diseases (Ell et al. 2014; Wang et al. 2017).

In light of a growing concern regarding the overinterpretation of small changes in gene expression often reported regarding mRNA targeting by miRNAs (Seitz 2017), the identification of the most potent large scale changes in gene regulation may be the most biologically relevant interactions. The potential questionable significance of modest changes in gene expression as a result of mRNA-miRNA interactions has been brought to question as a result that the variability observed in gene expression between individuals can overshadow these small changes (Pinzon et al. 2017). In order to identify the largest scale changes in protein expression upon the expression of the miR-23~24 cluster, we have focused on label-free quantitative proteomics using the model system of HEK 293T cells, which do not express the miR-23~24 cluster at detectable levels (Figure 1).

While isotopic labeling methods, such as iTRAQ isobaric tags (Wiese et al. 2007), are widely considered to be superior for quantitative proteomics, technical issues regarding their limitation in detecting large differences in abundance is often overlooked (Ow et al. 2009; Wang et al. 2012). The suppression of the differences quantified by these labeling methods is likely a result of contaminating co-eluting peptides into the collision cell of the mass spectrometer. While modern instruments have sufficient resolution to resolve nearly isobaric masses, they nonetheless suffer from poor resolving power for the isolation of parental ions for MS/MS fragmentation, which is typically conducted by quadrupole. As a result, multiple species may contaminate the collision cell and lead to contaminating reporter ions, an occurrence which will be more significant for low abundance peptides. This results in the compression of the data with lower
fold differences being observed between samples as has been previously reported with head-to-head comparisons of iTRAQ and label-free proteomic analysis (Latosinska et al. 2015; Trinh et al. 2013). For example, a previous study on the proteome changes upon miR-23–24 cluster expression reported a maximal ~2-fold change for the most significant protein alterations (Ludwig et al. 2016). In this previous work Dhfr and Nolc1 were two of the more significant changes observed upon expression of the miR-23–24 cluster, with reduced levels of protein of approximately 40 and 60% respectively. While these changes may be significant our label free quantification was unable to detect a significant reduction in these proteins, our data revealed a reduction of ~30% for both of these proteins by the variability between replicate analysis resulted in no statistically significant difference being observed (Supp. Table 1).

**Proteomic analysis**

While label-free analysis may be insensitive to the quantification of small changes in protein abundance, the analysis of replicate data in Figure 2A reveals sufficient reproducible quantification between analysis. Between these replicates the overall $r^2$ for all the data was 0.92. Nonetheless, the reproducibility between runs is not exact, with increasing variability for low abundance proteins, and highlights the need for replicate analysis for reliable label-free proteomic analysis.

The complete comparison of the proteomes of miR-23–24 cluster expressing, or vector control cells lead to the identification and quantification of 4,349 proteins (Supp. Table 1) that range in over five orders of magnitude in abundance as determined by their relative ion intensity (Figure 2). The comparison of protein abundance between the samples revealed a large-scale reduction in a large number of proteins, careful analysis of the volcano plot in Fig 2C, reveals a large number of proteins are observed to change in abundance by 10 to 100-fold, even if excluding those proteins uniquely observed in one experimental condition. This is in stark contrast to the differences reported by iTRAQ analysis (Ludwig et al. 2016) and demonstrates the utility of label-free proteomic analysis for the quantification of large changes to the proteome. The statistical comparison of the quantified differences in the proteomes of the miR-23–24 cluster expressing cells versus the vector control cells identified variations in large numbers of proteins at different levels of statistical confidence (Figure 2E).

With the large numbers of changes in protein abundance determined, ~750 at a p-value of <0.05 and ~200 at a p-value of <0.01 (Figure 2E), a simple interpretation of the molecular pathways regulated by the miR-23–24 cluster is not possible. Gene Ontology analysis of the proteins observed to change using a cut off of a p-value <0.01, revealed an enrichment of the five KEGG groups listed in Table 1. While this list includes ribosomes, peroxisomes, splicing and a few metabolic pathways, it is not entirely insightful regarding the specific molecular function of the miR-23–24 cluster, however, it is in agreement with previous reports linking the miRNAs from this cluster to metabolism (Gao et al. 2009; Kelly et al. 2015). Surprisingly, within the groups of protein populating the functions listed in Table 1, most are not identified or predicted targets of miR23a, miR27a or miR24-2 with the exception of a couple proteins like Rpl19 which is a predicted target of miR27a (Chou et al. 2018). This lack of association suggests many of the observed changes are a result of indirect effects of the activities of the miRNAs from the miR-23–24 cluster. The occurrence of the regulation of broad pathways by this cluster are not surprising given diverse impact of this miRNA cluster in cell differentiation (Cho et al. 2016; Kurkewich et al. 2017) and disease  (Bang et al. 2012; Ell et al. 2014; Wang et al. 2017).
To further query the top group identified in Table 1, ribosomes, we then went on to investigate the data obtained from all the ribosomal proteins detected and not just those that met the statistical threshold of \( p<0.01 \). This secondary analysis was revealing in that it is suggestive of a general global down regulation of ribosomal proteins with the exception of a few ribosomal proteins that are observed to increase in abundance (Figure 3). While the increase in abundance of select ribosomal proteins may at first be counter intuitive with regards to a model for a global down regulation in ribosome biogenesis, the individual increase in select ribosomal proteins may reflect their specific or alternative roles. For example, while RpsA is a known small ribosomal subunit protein (Malygin et al. 2011) it is also known to have extra-ribosomal functions as a laminin receptor (Scheiman et al. 2010) so its observed increase in abundance may be related to this function.

**Target Identification**

While proteomics analysis reveals the global changes to the proteome upon expression of the miR-23~24 cluster, it does not provide insight regarding which changes may be a direct result of miRNA targeting. Direct target identification and validation has demonstrated to be challenging, with current reports having to focus on one or two targets of a miRNA in a single study. For example, a recent report on miR27a has identified both interleukin-10 and TGF-β-activated protein kinase 1 binding protein as targets of this miRNA (Hussain et al. 2017). A challenge in the identification of miRNA targets is that prediction algorithms, such as those used by TargetScan (Agarwal et al. 2015), predict enormous number of candidates which need to be validated on a case by case scenario as the success rate for identification can not be reliably estimated as a result of a lack of negative results being published in the literature. Within our own data set the predictions almost appear random (Figure 4) and is in agreement with previous reports where minimal overlap of target prediction and proteomic changes are observed (Lee et al. 2015). When classifying the up and down regulated proteins observed upon miR-23~24 expression, the percent of predicted targets in both groups are similar. This is not to say that these prediction algorithms are not of high utility, but one must be cautious regarding the interpretation of pure bioinformatic analysis of miRNA targeting (Chhabra et al. 2010). The utility of mRNA target prediction algorithms and bioinformatic analysis are of greatest utility when combined with biological data to identify key target candidates. For example, this cross-platform analysis was demonstrated successfully with the identification of the miR27a-calreticulin signaling axis (Colangelo et al. 2016).

With our focus to identify the most potent mRNA targets of the miRNAs derived from the miR-23~24 cluster, our proteomics data was cross correlated with the miRNA targets from miRTarBase (Chou et al. 2018) and TargetScan (Agarwal et al. 2015). As mentioned, the overlap of predicted miRNA targets and our quantified proteomic changes is marginal (Figure 4), of the 448 down regulated proteins, at a p-value of <0.05, only 89 are predicted targets of either miR23a, miR27a or miR24-2. As all the alterations to the proteome are not likely to all be a result of direct regulation by miRNAs this finding is somewhat expected. In contrast to this, of the 307 proteins observed to increase in abundance upon the expression of the miR-23~24 cluster, 60 are predicted to be targets of one of the three miRNAs. At first glance the somewhat random occurrence of miRNA predicted targets and cellular changes to the proteome leads to what may be considered a poor result regarding the predictive power of identifying miRNA targets using these algorithms. Nonetheless, in combination with biological data the results are
still very promising, as the intersection of the datasets reveals the most promising candidate
genes targeted by the miRNAs from the miR-23–24 cluster.

In returning to the original question regarding the identification of candidate miRNA
targets that lead to large reduction in expression by a miRNA, our data was cross referenced with
the miRTarBase and TargetScan databases for miRNA targets to identify proteins that met the
following criteria. Proteins that either were reduced in abundance by at least ten-fold or were
only detected in the vector control cells, exhibited a p-value of <0.005 for the T-Test between the
control and miR-23–24 replicates and were predicted to be targets of one of the three miRNAs.
These very restrictive constraints lead to the identification of the 11 proteins listed in Figure 5.
While most of these proteins, such as Orc6 have not been previously reported to be associated
with any of the miRNAs from the cluster, grancalcin (GCA) has been reported previously to be
regulated by miR27a (Schoolmeesters et al. 2009). For future investigations on this cluster, these
proteins are top candidates for regulation. If the criteria are relaxed, to a p-value <0.01 for
example but keeping the criteria of the ten-fold change, the number of protein candidates jumps
to 25 (Supp.Table 3).

Concluding Remarks

Here we report on the utilization of label-free proteomics to characterize the global changes to
the cellular proteome upon the expression of the miR-23–24 cluster and reveal the utility of this
approach to identify the most promising potent targets of the miRNAs originating from this
cluster. While label-free proteomics is traditionally not viewed to be as quantitative and more
problematic than isotopic labeling techniques, such as iTraq, we demonstrate it is of high utility
in identifying the largest magnitude changes, which are desired for the identification of the most
potent targets for miRNA regulation. While our analysis to identifying the most significant
inhibition by miRNAs from the miR-23–24 cluster revealed a previously identified target, our
data reveals an additional series of previously unreported potential targets which exhibit reduced
expression by over an order of magnitude upon expression of the miR-23–24 cluster. In contrast
to previous proteomic investigations on this cluster using iTraq labeling which only reported on
changes of less than two-fold in protein abundance.

Acknowledgements

This project was supported by a Discovery Grant from the Natural Sciences and Engineering
Research Council (NSERC) of Canada to R.P.F. Conflict of interest statement: The authors
declare that they have no conflict of interest associated with this work.
References


cluster promotes breast cancer lung metastasis by targeting metastasis-suppressive gene


quantification on quadrupole-orbitrap mass spectrometer. Mol Cell Proteomics 11(12): 1709-

Gao, P., Tchernyshyov, I., Chang, T.C., Lee, Y.S., Kita, K., Ochi, T., et al. 2009. c-Myc suppression of miR-
23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. Nature

doi:10.1038/nature03552.


Antimicrobial Responses of Murine Macrophages Infected by Mycobacterium avium subspecies
paratuberculosis by Targeting Interleukin-10 and TGF-beta-Activated Protein Kinase 1 Binding

parameters affecting calcium-phosphate precipitate formation. Nucleic Acids Res 24(4): 596-

metabolism using miR-23 tips the balance towards a highly productive phenotype. Biotechnol J
Draft


https://mc06.manuscriptcentral.com/bcb-pubs


### Table 1. Pathway Enrichment for Proteins Determined to Change in Abundance at a Statistical Cut Off of <0.01.

<table>
<thead>
<tr>
<th>Pathways</th>
<th>Count</th>
<th>p-Value</th>
<th>KEGG</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ribosome</strong></td>
<td>RPLP0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MRPL12, MRPL14, MRPL3, RPL11, RPL19, RPL23, RPL32, RPL38, RPS11, RPS14, RPS6, RPS8</td>
<td>5.2E-07</td>
<td>hsa03010</td>
</tr>
<tr>
<td><strong>spliceosome</strong></td>
<td>PUF60,</td>
<td>6.6E-04</td>
<td>hsa03040</td>
</tr>
<tr>
<td></td>
<td>LSM3, LSM7, RBM8A, PPIH, SRSF10, SRSF9, SNRPF, ZMAT2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RNA degradation</strong></td>
<td>ENO1, PNPT1</td>
<td>0.025</td>
<td>hsa03018</td>
</tr>
<tr>
<td></td>
<td>LSM3, LSM7, MPHOSPH6</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>peroxisome</strong></td>
<td>NUDT19</td>
<td>0.032</td>
<td>hsa04146</td>
</tr>
<tr>
<td></td>
<td>DECR2, PEX11B, SCP2, SOD1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Glutathione</strong></td>
<td>-</td>
<td>0.038</td>
<td>hsa00480</td>
</tr>
<tr>
<td>metabolism**</td>
<td>GSTM2, GSTM3, GSTM4, MGST1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Oxidative</strong></td>
<td>-</td>
<td>0.044</td>
<td>hsa00190</td>
</tr>
<tr>
<td>phosphorylation**</td>
<td>ATP5F1, ATP5L, NDUFB6, NDUFB8, NDUFS5, UQCRB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1. miR-23~24 cluster expression. A) Northern Blot analysis for the expression of miR-23a in HEK 293T cells transfected with a plasmid to express the miR-23~24 cluster or the vector control. Total RNA from the indicated cell lines were included for comparison of endogenous miR23a levels in these cell types.

Figure 2. Label free proteomic analysis of miR-23~24 cluster expressing cells. A) The relative ion intensities for identified proteins from two biological replicates of control samples were plotted against each other to demonstrate the reproducibility of quantification and detection. B) The average relative ion intensities for proteins identified in the triplicate analysis of miR-23~24 expressing or vector control cells were plotted to demonstrate the global changes observed in the cellular proteomes. C) Volcano Plot of the data described in C), which includes the p-values determined for each protein quantified in the comparison of miR-23~24 expressing or vector control cells. Selected proteins from the data set are individually labeled. D) Quantified data from selected proteins are plotted to demonstrate the variability between experimental conditions. Major control proteins do not exhibit significant differences between the cells types (dark grey-miR-23~24 expressing cells: light grey-Vector control cells), while H2AFX, a known target of miR24a exhibits a nearly 10 fold reduction upon the expression of the miR-23~24 cluster. E) Summation of the number of changes in protein expression observed at the indicated statistical cut-offs.

Figure 3. Global Analysis and Selected Validation of Ribosomal Proteins. A) Analysis of all ribosomal proteins detected reveals an overall reduction in both small (left) and large (right) ribosomal sub unit proteins upon expression of the miR-23~24 cluster. In contrast to the global reduction in ribosomal proteins, the indicted ribosomal proteins exhibited increased abundance upon miRNA cluster expression. B) Validation of the of the large fold changes observed upon miR-23~24 cluster expression with the initial shotgun proteomic analysis by parallel reaction monitoring of the indicated proteins.
Figure 4. Overlap of Predicted miR-23~24 Targets with Observed Proteomic Changes. Dark grey bars indicate the number of proteins identified to be targets of miR23a, miR27a or miR24-2 in either Target Scan or miRTarBase. Light grey bars indicate the number of proteins observed to change in abundance but are not predicted targets of any of the three miRNAs. The change in whether protein abundance increased or decreased in abundance upon miR-23~24 expression is indicated. The changes in abundance are grouped according to the statistical confidence of the change in abundance observed. Of note is the lack of a significant correlation of proteins predicted to be targets of the miRNAs and them decreasing in abundance upon miRNA cluster expression.

Figure 5. Top Candidate Target Genes for the miR-23~24 Cluster. The top candidate potential direct targets of the miRNAs from the miR-23~24 cluster were identified by those were only detected in the control cells or exhibited a greater than 10-fold reduction upon miR-23~24 expression. To minimize this list to the 11 proteins shown, only those with a p-value of under 0.005 are displayed. ** indicate that the protein was not detected in the miR-23~24 cluster expressing cells.
Figure 1. miR-23~24 cluster expression. A) Northern Blot analysis for the expression of miR-23a in HEK 293T cells transfected with a plasmid to express the miR-23~24 cluster or the vector control. Total RNA from the indicated cell lines were included for comparison of endogenous miR23a levels in these cell types.
Figure 2. Label free proteomic analysis of miR-23~24 cluster expressing cells. A) The relative ion intensities for identified proteins from two biological replicates of control samples were plotted against each other to demonstrate the reproducibility of quantification and detection. B) The average relative ion intensities for proteins identified in the triplicate analysis of miR-23~24 expressing or vector control cells were plotted to demonstrate the global changes observed in the cellular proteomes. C) Volcano Plot of the data described in C), which includes the p-values determined for each protein quantified in the comparison of miR-23~24 expressing or vector control cells. Selected proteins from the data set are individually labeled. D) Quantified data from selected proteins are plotted to demonstrate the variability between experimental conditions. Major control proteins do not exhibit significant differences between the cells types (dark grey-miR-23~24 expressing cells: light grey-Vector control cells), while H2AFX, a known target of miR24a exhibits a nearly 10 fold reduction upon the expression of the miR-23~24 cluster. E) Summation of the number of changes in protein expression observed at the indicated statistical cut-offs.

745x540mm (96 x 96 DPI)
Figure 3. Global Analysis and Selected Validation of Ribosomal Proteins. A) Analysis of all ribosomal proteins detected reveals an overall reduction in both small (left) and large (right) ribosomal subunit proteins upon expression of the miR-23~24 cluster. In contrast to the global reduction in ribosomal proteins, the indicated ribosomal proteins exhibited increased abundance upon miRNA cluster expression. B) Validation of the large fold changes observed upon miR-23~24 cluster expression with the initial shotgun proteomic analysis by parallel reaction monitoring of the indicated proteins.
Figure 4. Overlap of Predicted miR-23~24 Targets with Observed Proteomic Changes. Dark grey bars indicate the number of proteins identified to be targets of miR23a, miR27a or miR24-2 in either Target Scan or miRTarBase. Light grey bars indicate the number of proteins observed to change in abundance but are not predicted targets of any of the three miRNAs. The change in whether protein abundance increased or decreased in abundance upon miR-23~24 expression is indicated. The changes in abundance are grouped according to the statistical confidence of the change in abundance observed. Of note is the lack of a significant correlation of proteins predicted to be targets of the miRNAs and them decreasing in abundance upon miRNA cluster expression.
Figure 5. Top Candidate Target Genes for the miR-23~24 Cluster. The top candidate potential direct targets of the miRNAs from the miR-23~24 cluster were identified by those were only detected in the control cells or exhibited a greater than 10-fold reduction upon miR-23~24 expression. To minimize this list to the 11 proteins shown, only those with a p-value of under 0.005 are displayed. ** indicate that the protein was not detected in the miR-23~24 cluster expressing cells.