2021

Identification and Optimization of miRNA Biomarkers for Body Mass Index (BMI) Estimation in Forensic Samples

Jennifer J. Shealy
Virginia Commonwealth University

Identification and Optimization of miRNA Biomarkers for Body Mass Index (BMI) Estimation in Forensic Samples

Jennifer Shealy
Fall 2020 – Spring 2021

April 12th 2021

Director: Sarah J. Seashols-Williams, Ph.D.
Williams Research Laboratory
Department of Forensic Science
Virginia Commonwealth University

Acknowledgements

I would like to thank my research advisor, Dr. Sarah Seashols-Williams, along with my other committee members, Dr. Edward Boone and Dr. Catherine Connon, for allowing me to complete this project and for their support and expertise throughout. I would also like to thank the members of the Williams and Dawson-Green labs for all of their help and answered questions while completing this project.

Abstract

DNA profiles do not provide as much evidentiary value when either all known samples can be excluded or when there is no known suspect. DNA phenotyping is a technique used for investigative purposes to predict externally visible characteristics (EVCs) based on specific markers found in the genome. Most phenotyping panels are currently restricted to markers stable over a lifetime within an individual and are not able to predict environmental or metabolic impacts on EVCs. It is known that miRNA expression levels change due to environmental and metabolic factors, such as BMI, and research has proven associations between circulating plasma miRNAs and BMI. In this project, 25 dried whole blood swabs were prepared from individuals with varying BMI values. To accommodate the typical forensic casework protocol, a DNA extraction method (QIAamp DNA Investigator Kit) was utilized followed by a cDNA synthesis reaction. Fifteen candidate miRNAs were examined for their expression levels and analyzed against both BMI and weight. It was found that the calculated change in quantification cycle from stable miRNA expression, or ΔCq, of miR-486-5p and miR-885-5p both individually showed negative associations with BMI. The ΔCq of miR-486-5p additionally showed a negative association with weight, along with the ΔCq of let-7i-5p and the Cq of miR-194-5p. The observed associations were found to be weak, but it is proposed that this is mostly due to the small sample size of this study. However, the data collected in this study, when analyzed using predictive models, was shown to have some success with a classification and regression tree analysis and a high level of success when utilizing a support vector machine model. This research demonstrates the possibility of adding environmentally impacted EVCs into current phenotyping panels. Recommendations for future work include the testing of additional markers and the use of a larger set of samples from the population.

Keywords: forensic science, forensic biology, DNA phenotyping, BMI estimation, blood swabs, miRNA expression

Introduction

The use of short tandem repeats (STRs) found in the human genome for forensic analysis has been regarded as the gold standard in the field of forensic biology. These have proven to produce reliable results even with low-level and partially degraded DNA, often the case of forensic samples (1). While this type of analysis can be imperative in the understanding of a crime, the DNA profile produced may only be useful in the comparison to known profiles. When suspects are eliminated from contributing to the DNA profile found on evidence or when there are no known suspects nor databank hits, further examination may give investigators leads based upon phenotypic features inferred from the same biological evidence.

Single nucleotide polymorphisms (SNPs) are known to be useful identifiers for specific genes and have been used as markers to predict various phenotypic appearances (2). SNPs associated with phenotypes due to pigmentation have been some of the most investigated. For example, blue versus brown eye color has been shown to be highly predictive with about a 90% success rate for each using the IrisPlex multiplex assay (3,4). In addition, the newer HIrisPlex system combines eye color prediction with reliable hair color prediction, both independent from bio-geographic ancestry (5). While less reliable and much more explorative, SNPs have been recognized as potential biomarkers for the prediction of facial features as well, such as eyebrow width, eye distance. In addition, previously identified markers for sex and biogeographic ancestry have also been found to aid in the prediction of facial features (6). Methods outside of using SNPs have also been identified as predictors of phenotypes such as DNA methylation. certain DNA methylation sites are associated with the age of an individual and can accurately predict age (±6 years) 77.30% of the time in individuals under 60 years of age (7). Known age-

informative CpGs have since been investigated as biomarkers with low-level DNA input, as the case with most forensic samples, showing promising results with as low as 20ng of DNA (8).

MicroRNAs (miRNAs), particularly extracellular miRNAs, are RNA molecules 18-24 nucleotides long and are known to have regulatory functions with neighboring cells (9). Their regulatory functions and stability due to their small size have proven miRNAs to be useful in the determination of the age of a biological stain as well as the age of an individual (10,11). Similar to age of individuals, miRNA expression levels have been shown to suggest postmortem intervals in incisional wounds (12).

Body mass index (BMI) is a measurement of body fat, and is calculated by dividing an individual's weight (kg) by their height squared (m^2). BMI or body weight prediction would provide an immense asset to the forensic community if added to a larger panel of phenotype markers, providing another piece of information to create a more accurate image of an individual solely from a biological sample. While not yet investigated for forensic purposes, there have been significant advances in biomedical research regarding the relationship of certain biomarkers and BMI due to the known correlations that BMI has with various health risks. Similar to the findings about age and the methylation of CpG sites (7), methylation levels at certain loci in blood cells have been found to be associated with BMI (13). It has been found that high BMI levels are also associated in changes of messenger RNA (mRNA) expression, including inflammatory related mRNAs in the brain (14). Similarly, certain variants in BMI-related genes affect the efficiency of miRNAs to bind to those genes, making them less able to block the production of protein (15).

As the understanding of miRNAs and their functions are explored, so are the relationships between expression levels and phenotypes. As the body mass index (BMI) of an

individual can be highly dependent on metabolic factors, biomarkers such as SNPs become less useful in the estimation of BMI from body fluid samples. miRNAs become an ideal candidate for predicting this category of phenotypes as the regulation of genes change alongside these metabolic factors.

Specific miRNA candidates that correlate with BMI have been identified in published biomedical research, demonstrating linear relationships with miRNA expression levels in plasma and a varying range of BMIs. In a study investigating predictors of metabolic syndrome, relationships were observed between two circulating plasma miRNAs and BMI, as well as relationships with other predictors such as waist circumference, blood pressure, and plasma glucose (16). A total of 19 circulating plasma miRNAs were found to have significant associations with BMI, with three specific miRNAs (miR-122-5p, miR-148a-3p, miR-505-3p) showing the most significant associations with $q < 0.001$ (17). In addition, some of those identified miRNA markers have been confirmed to correlate with BMI through previous studies in our research laboratory using RNA from plasma samples from individuals with known BMIs ranging from normal weight (BMI = 18.5-24.9) to obese (BMI ≥ 30), with the most interesting trends seeming to show between BMI and miRNAs miR-106a, miR-185-5p, and let-7i-5p (**Figure 1**). It is hypothesized that these same correlations between miRNA expression levels and BMI are to be observed in dried whole blood, commonly encountered in forensic evidence samples, along with various other body fluids which would allow for the addition of these biomarkers to a larger phenotypic panel to aid in investigative leads.

Furthermore, it has been previously demonstrated in our laboratory that traditional forensic DNA extraction methods are consistently effective in coextracting miRNA with DNA, although a lower miRNA yield than that observed with RNA specific extraction methods was

observed (18). It is therefore additionally hypothesized that using a DNA specific extraction method in this study would be efficiently isolate miRNA for downstream expression profile analysis, indicating the applicability of integrating miRNA-based techniques into forensic biology workflows.

Methods

Sample Selection and DNA/RNA Extraction

A total of 25 dried whole blood samples of known origin collected on sterile cotton swabs and stored at room temperature were taken from the VCU Forensic Science Department IRB-approved Forensic Science Biological Samples Registry (HM20002931). BMI was calculated and categorized according to the Centers for Disease Control and Prevention guidelines (19). Samples were selected so that all BMI categories were represented, ranging from underweight to obese, as shown in **Table 1**. Based on the demonstrated ability to isolate miRNA from DNA extracts (17), DNA was extracted from the samples using the QIAcube (Qiagen, Hilden, Germany) instrument using the QIAamp DNA Investigator Kit reagents and following the *Forensic Casework* sample protocol with an elution volume of 30µL. Extracts were stored at -20°C until use.

Reverse Transcription of miRNA Extracts

Reverse transcription was performed on the Proflex PCR System (Thermo Fisher Scientific, Waltham, MA) using the qScript™ miRNA Quantification System (Quanta Biosciences, Gaithersburg, MD). Using the reagents provided from the the qScript™ miRNA Quantification System (Quanta Biosciences), the first reaction was performed by adding 2µL Poly(A) Tailing Buffer, 7µL of extracted DNA, and 0.6µL of Poly(A) Polymerase, followed by

an incubation of 37°C for 40 minutes, and then 70°C for five minutes. The cDNA synthesis reaction was performed by adding 9µL of microRNA cDNA Reaction Mix and 1µL of qScript Reverse Transcriptase to the previous reaction mix. This reaction was incubated for 20 minutes at 42°C followed by five minutes at 85°C. The resulting cDNA solution was stored at -20°C until use.

Amplification and Detection

Specific qPCR primers were used for each candidate miRNA and obtained from IDT (Integrated DNA Technologies, Coralville, IA, USA) (**Table 2**). All new primers were tested prior to the experiment using samples consisting of DNA extract, DNase treated DNA extract, reverse transcribed miRNA from DNA extract, and reverse transcribed miRNA from DNase treated DNA extract to ensure specificity of the primer to miRNA. Only primers that showed similar amplification of product in the reverse transcribed DNA extract sample and product in the reverse transcribed DNase treated extract without amplification in either of the two non-reverse transcribed DNA extract samples were used.

Each miRNA target was amplified in single wells for each sample by mixing 6.25µL PerfeCTa SYBR Green SuperMix (2X) (Quanta Biosciences), 0.25µL PerfeCTa microRNA Assay Primer (10µM) (Integrated DNA Technologies), 0.25µL PerfeCTa Universal PCR Primer (10µM) (qScript™ miRNA Quantification System, Quanta Biosciences), 3.75µL nuclease-free water, and 2µL microRNA cDNA. The 3-step cycling protocol was as follows: 2min pre-incubation at 95°C; 40 PCR cycles consisting of 95°C for 5sec, 60°C for 15sec, and 70°C for 34sec; and a final extension stage comprised of 95°C for 15sec, 60°C for 1min, and 95°C for 15sec.

Raw data was analyzed within the QuantStudio ™ Real-Time PCR software v1.3 (Thermo Fisher Scientific) using a quantification cycle (Cq) threshold of 0.015. Differential expression or delta cycle threshold (ΔCq) values were calculated by subtracting the candidate miRNA Cq value from the let-7g Cq value, as let-7g has been shown to show stable expression across samples (20). All samples with an undetectable amount of miRNA expression were given a Cq value of 40 for analysis purposes.

Data Analysis

Data was first visualized using Microsoft Excel (Microsoft Corporation, Redmond, WA) to show general trends between BMI, ΔCq, Cq, and weight. Promising markers from visual examination of the data were then analyzed in RStudio v4.0.2 (Integrated Development Environment for R, Boston, MA) (21). Linearity and homoscedasticity were confirmed for all sample sets using quantile-quantile plots. A simple linear regression test was performed; due to the preliminary nature of this work, a confidence level of 0.90 was utilized ($p < 0.10$ was considered statistically significant). To assess the predictive qualities and strength of the miRNA markers selected in this study, data from all markers (both raw Cq and ΔCq) were combined and used in a classification and regression tree (CART) analysis in RStudio using the rpart and rpart.plot packages. Additionally, a support vector machine (SVM) model (type: C-classification; kernel: linear; cost: 1) was analyzed using the package e1071. R code for each analysis is shown in **Supplemental Table 1**.

Results and Discussion

Sample Selection and Primer Validation

Selection of samples used in this project were limited to those with recorded weight and height. The CDC's guidelines for BMI calculation were used to determine if the samples belonged to the underweight, normal, overweight, or obese categories. Samples were then chosen in order to have representation among all four categories, as shown in **Table 1**. Due to the limitations of only being able to use samples with recorded weight and height and not being able to collect new samples due to the COVID-19 pandemic, all BMI categories could not be equally represented. In addition, two miRNA primers selected for this study could not be validated using the methods described above. These two miRNA candidates, miR-193b-3p and miR-122-5p, were therefore not tested or further used for this study.

Preliminary Analysis/Visualization

Analysis of the raw results from amplification of the selected miRNA candidates showed that seven of the 15 targets had at least one sample with undetectable amounts of miRNA expression. While it has been demonstrated that miRNA can be coextracted using traditional DNA extraction methods, there is known reduction in sensitivity when comparing RNA extraction methods to DNA extraction methods for the purpose of extraction RNA (19). While this issue could be resolved using specific RNA extraction techniques, it would be at the cost of implementing a separate step from the already established DNA workflow for the lab analyst, costing time, materials, and consumption of sample. It is important to note that while random undetectable miRNA expression is an issue with some of the observed results of this experiment, observing an entire category of samples with mostly undetectable miRNA expression may strengthen the power of using these miRNAs to discriminate between BMI categories.

When amplification results were analyzed visually, there was also interest in evaluating the data strictly by weight as well as the initial proposal of analyzing BMI, as it is known that

BMI is not a perfect system of determining body fat percentage (22). Two specific miRNA targets, miR-486-5p and miR-885-5p, showed interesting trends when analyzed using scatterplots based on ΔCq and BMI (**Figure 2**). Additionally, ΔCq expression from miR-486-5p seemed to show a linear trend when compared against donor weight (**Figure 3**). Two final markers of interest were let-7i-5p and miR-194-5p, which seemed to show trends between their respective raw Cq values and donor weight (**Figure 4**). In contrast, miR-143-3p demonstrated no apparent trend in the data and therefore was not selected for any further statistical analysis (**Figure 5**). Once these preliminary visual identifications were made, miR-486-5p, miR-885-5p, miR-486-5p, let-7i-5p, and miR-194-5p were analyzed for association using RStudio. All selected targets were firstly analyzed using quantile-quantile plots to determine linearity and homoscedasticity (**Figure 6**).

Association Analyses

When preliminarily visualizing data, a seemingly linear relationship was observed between ΔCq and BMI for both miR-486-5p and miR-885-5p (**Figure 2**). The results from a simple linear regression analysis showed that the ΔCq value of miR-486-5p ($p = 0.0822$, $R^2 = 0.1256$) and miR-885-5p ($p = 0.0939$, $R^2 = 0.1172$) could explain 12.56% and 11.72% of the observed variability in BMI respectively, leading to a determination that both have a negative association with BMI (**Figure 7A&B**). Likewise, analysis of the ΔCq from miR-486-5p ($p = 0.036$, $R^2 = 0.1775$) showed it could explain 17.75% of the observed variability in donor weight, demonstrating a negative association as well (**Figure 7C**).

While the previously mentioned associations were found using ΔCq, let-7i-5p and miR-194-5p seemed to show a linear relationship between the raw Cq values and donor weight (**Figure 4**). Using a simple linear regression analysis, let-7i-5p ($p = 0.04977$, $R^2 = 0.1572$) and

miR-194-5p (p = 0.0866, R^2 = 0.1223) could explain 15.72% and 12.23% of the observed variability in donor weight respectively, leading to a determination that both have a negative association with donor weight (**Figure 8**).

Markers found to be significant from the simple linear regression analyses with similar predictors were analyzed together using a multiple linear regression analysis. When analyzed together against the calculated BMI, the ΔCq values from miR-486-4p and miR-885-5p did not prove to hold any statistical significance (p = 0.1085). Additionally, the Cq values from let-7i-5p and miR-194-5p were analyzed together using a multiple linear regression analysis, but had no statistical significance as well (p = 0.1518).

While not found to be statistically significant, interesting trends were observed for several miRNA targets from the visual examination of the data by scatterplots. Interesting trends of note were between ΔCq and BMI with two observed outliers for miR-365-3p, between ΔCq and BMI with three observed outliers for miR-20a-5p, between ΔCq and weight for miR-145-5p, and between Cq and weight for miR-505-3p when excluding observed outliers from undetectable samples (**Figure 9**).

While it was initially proposed that the trends observed for circulating plasma miRNAs would continue with dried whole blood samples, starkly different trends were observed. For example, when using blood serum and an RNA extraction method, no meaningful trends were observed when ΔCq was plotted against BMI for either miR-486-5p or miR-885-5p (**Figure 1**). However, these two markers showed a statistically significant association between ΔCq and BMI with dried whole blood samples and using a DNA extraction method, as discussed above. Additionally, markers which seemed to be promising from the previous validation, such as miR-106a or miR-99a-5p, did not show any significance or possible trends in this study. It should also

be noted that the previous validation using blood serum and an RNA extraction method did not include any underweight individuals. Even so, the visualized trends from the normal to obese samples did not correspond with the results presented from this study either.

The most obvious potential explanation for the differences observed in this study could be due to the shift from analyzing total miRNA from blood serum to total miRNA in whole blood. While individual miRNA expression profiles may be consistent when analyzing circulating miRNAs specifically, analyzing miRNA expression from whole blood will also display any intracellular miRNA expression from leukocytes and erythrocytes as well, potentially changing the entire expression profile (23, 24). This is especially true for this study and miRNAs associated with BMI, as leukocyte count can increase with higher BMI levels (25). In addition to leukocytes, erythrocytes also contain miRNA and miR-486-5p, a statistically significant marker in this study, has specifically been observed to be one of the major miRNAs found in erythrocytes (26). It is also a possibility that the chosen extraction method could be a factor in the difference in observed results. Random samples with undetectable amounts of amplification product were observed for some miRNA targets in this study, and therefore contributed as outliers, potentially skewing trends with either BMI or donor weight.

For the four markers found to hold statistically significant associations from the simple linear regression analyses, boxplots were constructed to analyze any potential differences observed between the reported sexes of the donors. Plots were first made to compare if any overall differences between sexes for both BMI and donor weight could be identified (**Figure 10**). As it was already known that more female samples were used in the sample set for this study (19 females, six males), it was not surprising to see that females were shown to have a lower average weight than males, as males were more underrepresented in the underweight and normal

BMI categories, however distribution of BMI was similar between the two. Plots were also constructed to show the distribution of either ΔCq or Cq for each marker found to be of statistical significance from the simple linear regression analysis with either BMI or donor weight (**Figure 11**). Although differences were observed, distributions were not different enough to be unexpected considering the sample size. While no notable differences were observed in regards to sex in these results, future studies should try to maintain equal representation of sexes within sample sets when available.

It is not too surprising that the four significant markers identified from the simple linear regression analyses in this study have some sort of association with weight or the calculated BMI as there have been published studies discussing their potential roles with weight-related factors. The first discussed target, miR-486-5p, has been demonstrated to be involved in the regulation of proliferation of human adipose-derived mesenchymal stem cells (27). While this does not explain the observed results from this study, knowing the roles that these miRNAs play with adipocytes is not surprising considering the relationship between adipocyte hyperplasia and BMI (28). In addition, an upregulation of miR-885-5p has been observed in individuals with fatty liver disease (29). The association observed in this study may potentially be partially explained by this association as BMI is known to be associated with fatty liver risk (30). Another significant marker, let-7i-5p, has been suggested to be involved in the browning process of white adipocytes, having a decreased expression in brite adipocytes (31). While more studies need to be performed to confirm, these findings seem to contradict the results observed from this study, since it can be inferred that individuals with lower overall weight would have increased browning of adipose tissue when compared to individuals with higher weight; however, the overexpression of let-7i-5p was observed to repress brite adipocyte function (31). Lastly, miR-

194-5p is not very well characterized with its relationship to BMI or weight; however, it has been observed that there is a significant decrease in expression levels when analyzing the kidney tissue of diabetic mice (32).

Predictive Models

Although any associations analyzed were found to be weak, a CART analysis was included to determine when Cq and ΔCq data from all markers were combined, if any useful determinations could be made to allow for the classification of samples into BMI groups. The resulting classification tree is shown in **Figure 12A** with the Cq data from miR-505-3p being recognized as the most predictive, separating samples with a Cq value ≥ 35 into the normal weight category, with the rest of the samples categorized as overweight. As shown in **Table 3A**, eight out of the nine normal weight individuals were correctly classified (88%), while one individual was misclassified as overweight. Five out of seven individuals were correctly classified as overweight (71%), with two individuals misclassified as normal. Four of the obese individuals were classified as normal, two obese individuals classified as overweight, and all underweight individuals were classified as normal (52% correct overall). The same results were yielded when a CART analysis was performed using strictly raw Cq data.

Additionally, a CART analysis was performed using strictly ΔCq values for each target miRNA, and miR-194-5p was determined to be the most predictive separating individuals with a ΔCq of -4.6 or greater into the normal category, and all other individuals into the obese category (**Figure 12B**). As shown in **Table 3B**, eight normal weight individuals were correctly classified (88%), while one individual was classified as obese. Four out of the five obese individuals were correctly classified (80%), with one individual incorrectly classified as normal. Five overweight individuals were classified as normal, with the other two overweight individuals classified as

obese. One underweight individual was classified as normal, with the other four classified as obese (48% correct overall). The addition of sex as a predictor in all CART analyses did not prove to contribute to any predictions.

In addition to the CART analyses, a SVM model was used for classification prediction of individuals. The SVM was able to correctly classify each individual to its corresponding BMI category when blindly tested using both the raw Cq and ΔCq values collected from this study as a training set (100% correct overall) (**Table 4A**). While the correct classification rate for this model is incredibly high, it is important to emphasize the role that the small size of the dataset may play with these predictions. The small sample size contributes to the potential that the data is overfit, meaning more data will most likely render the model less accurate than observed in this study. Consequently, when only using raw Cq values as the training set, one individual was misclassified (94% correct overall) (**Table 4B**). A larger sample set will need to be analyzed to evaluate the true classification accuracy of using SVM as a model for phenotypic prediction.

Integration into Phenotypic Panel

Although the association between the analyzed miRNA expression patterns and BMI/weight were not considerably strong, there was some found success observed in the categorization of samples into BMI categories using a CART analysis and SVM model. These observed associations, along with the success in using this small sample set with predictive models, provides confidence that further research may discover that BMI or weight prediction in a phenotypic panel is very much possible. Further research with an expanded sample set including all BMI categories and a variance in weight of individuals is needed to confirm the associations and success with this project.

Along with an expanded sample set, an increase in miRNA targets is suggested as well, including miR-223, miR-125b, miR-221, miR-199b-3p, miR-199-5p, miR-100, miR-10a, miR-34a, miR-199a-5p, miR-1229, miR-210, miR-99b, miR-92a, miR-142-5p, miR-130b, miR-484, miR-133b, and miR-236. It has been observed that circulating miR-223 expression in blood serum is lower in individuals with higher BMI than in normal-weight individuals, and additionally has been shown to increase when the BMI of individuals is lowered due to lifestyle changes (33). Additionally, miR-133b and miR-236 should be investigated further as potential biomarkers to help identify underweight individuals. These two miRNAs have been observed to offer a potential connection between their expression values, nutrition, and age-related skeletal muscle decline when analyzing expression values in plasma (34). While a classification of underweight with respect to BMI does not strictly mean an individual is malnourished, it is proposed that these two miRNAs may potentially show an association between underweight individuals and observed expression. Lastly, the remaining miRNAs addressed above have been observed to have significant associations with BMI when analyzing subcutaneous adipose tissue (35). Since two of the associated markers in this study have been demonstrated to be expressed in and have functions with adipocytes, it is hypothesized that these listed markers may also have detectable levels of expression in whole blood samples. Additions to the four miRNAs observed in this study would be helpful in strengthening the value of BMI/weight predictive models when integrating into a phenotypic panel.

Expansion to Other Forensically Relevant Body Fluids

If continued associations were to be observed using an expanded sample set with dried whole blood, it is suggested that this project should be expanded to investigate associations with other forensically relevant body fluids, such as saliva. Since inflammatory response proteins can

be detected in saliva samples (36), it is proposed that saliva would potentially uphold many associations found between miRNA and BMI/weight observed in whole blood. It has been documented that adipose tissue is responsible for an inflammatory response to obesity (37). Since two of the miRNAs discussed in this study are known to be related to adipocytes and therefore potentially inflammation, inflammatory-related miRNAs (triggered by a high BMI) may potentially be able to be detected in saliva samples at similar expression patterns as shown in whole blood.

Conclusion

With increasing amounts of biological samples being submitted to forensic laboratories, the issue of producing DNA profiles without any useful known profile for comparison is also increasing. Predictive phenotyping is currently being introduced in such cases where a suspect may be unknown; however, this method for investigative leads currently only provides specific information from markers that are stable in the genome from birth. If associations between miRNA expression levels and metabolic or environmentally impacted phenotypes could be identified, characteristics such as BMI or weight could then be predicted just as hair, eye, or skin color. The goal of this project was to identify miRNA candidates that correlate with BMI using dried whole blood samples and methodology typical for a forensic laboratory to be used in further research for incorporation of these markers into a larger phenotyping panel for investigative purposes.

Previous work had identified multiple circulating plasma miRNAs shown to have strong associations with BMI (17), which were the markers selected for investigation in this study. In

contrast to the previous study, a DNA extraction method was used to show the ability to integrate miRNA analysis easily into the current typical forensic case workflow.

It was initially proposed that similar trends would be observed in this study as compared to those from a previous validation of the same markers using blood serum and an RNA extraction technique. The results of this study prove that while the same trends are not observed, miRNA targets analyzed still demonstrated statistically significant associations with either BMI or weight. Two candidate miRNAs (miR-486-5 and miR-885-5p) have an association with BMI and three candidate miRNAs (miR-486-5p, let-7i-5p, and miR-194-5p) have an association with weight. While these associations are not strong, these findings, put into perspective with the small sample size used, are promising for further research.

Through the use of a CART analysis as a means of a predictive model, eight out of nine normal weight individuals and five out of seven overweight individuals are correctly classified, with a total of 13 out of 25 total samples correctly classified using the combined Cq and ΔCq values. Using only raw Cq values, eight out of nine normal weight individuals and four out of five obese individuals are correctly classified, with a total of 12 out of 25 total samples correctly classified. All samples are correctly predicted into their respective BMI categories using an SVM model and combined Cq and ΔCq data. However, the possibility of the data being overfit due to the small sample size is very likely. This demonstrates the need for the addition of a larger sample size and the addition of more miRNA markers to further the ability to confidently separate between BMI groups.

In conclusion, the findings of this study prove that associations between miRNA expression due to a response to metabolic factors, such as BMI or weight, can be detected in complex forensic sample types such as whole dried blood. It is proposed that with a larger

sample set, stronger associations may be identified and, therefore, be used for the optimization of predictive models for either BMI or weight. In order to further this research, a larger sample size should be used along with the testing of more target miRNAs. If continued associations were to be found, predictive modeling applications, such as the CART analyses and SVM model predictions demonstrated in this study, could be optimized. If accurate, miRNA markers could be used in conjunction together for BMI or weight prediction and added to a larger forensic DNA phenotyping panel to be used for investigative purposes.

References

1. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B. Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. J Forensic Sci 2001;46(3):647-60.
2. Kayser M, De Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 2011(3):179-92.
3. Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int: Genet 2011;5(3):170-80.
4. Bulbul O, Zorlu T, Filoglu G. Prediction of human eye colour using highly informative phenotype SNPs (PISNPs). Aust J Forensic Sci 2020;52(1):27-37.
5. Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, Branicki W, Kayser M. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int: Genet 2013;7(1):98-115.
6. Serrano AC. Forensic DNA phenotyping: A promising tool to aid forensic investigation. Current situation. Span J Leg Med 2020;46(4):183-90
7. Park JL, Kim JH, Seo E, Bae DH, Kim SY, Lee HC, Woo KM, Kim YS. Identification and evaluation of age-correlated DNA methylation markers for forensic use. Forensic Sci Int: Genet 2016;23:64-70.
8. Heidegger A, Xavier C, Niederstätter H, de la Puente M, Pośpiech E, Pisarek A, Kayser M, Branicki W, Parson W, VISAGE Consortium. Development and optimization of the

VISAGE basic prototype tool for forensic age estimation. Forensic Sci Int: Genet 2020;48:102322.

9. Xavier C, de la Puente M, Mosquera-Miguel A, Freire-Aradas A, Kalamara V, Vidaki A, Gross TE, Revoir A, Pośpiech E, Kartasińska E, Spólnicka M. Development and validation of the VISAGE AmpliSeq Basic Tool to predict appearance and ancestry from DNA. Forensic Sci Int: Genet 2020;48:102336.
10. Zhao C, Sun X, Li L. Biogenesis and function of extracellular miRNAs. ExRNA. 2019;1(1):1-9.
11. Alshehhi S, Haddrill PR. Estimating time since deposition using quantification of RNA degradation in body fluid-specific markers. Forensic Sci Int 2019;298:58-63.
12. Fang C, Liu X, Zhao J, Xie B, Qian J, Li B, Zhang X, Wu H, Yan J. Age estimation using bloodstain miRNAs based on massive parallel sequencing and machine learning: A pilot study. Forensic Sci Int: Genet 2020;47:102300.
13. Ibrahim SF, Ali MM, Basyouni H, Rashed LA, Amer EA, Abd El-Kareem D. Histological and miRNAs postmortem changes in incisional wound. Egypt J Forensic Sci 2019;9(1):1-6.
14. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman ÅK, Aslibekyan S, Demerath EW, Guan W, Zhi D, Yao C, Huan T. Association of body mass index with DNA methylation and gene expression in blood cells and relations to cardiometabolic disease: a Mendelian randomization approach. PLoS Med 2017;14(1):e1002215.
15. Lauridsen JK, Olesen RH, Vendelbo J, Hyde TM, Kleinman JE, Bibby BM, Brock B, Rungby J, Larsen A. High BMI levels associate with reduced mRNA expression of IL10

and increased mRNA expression of iNOS (NOS2) in human frontal cortex. Transl Psychiatry 2017;7(2):e1044-.

16. Ramzan F, D'Souza RF, Durainayagam BR, Milan AM, Markworth JF, Miranda-Soberanis V, Sequeira IR, Roy NC, Poppitt SD, Mitchell CJ, Cameron-Smith D. Circulatory miRNA biomarkers of metabolic syndrome. Acta diabetol 2020;57(2):203-14.
17. Ameling S, Kacprowski T, Chilukoti RK, Malsch C, Liebscher V, Suhre K, Pietzner M, Friedrich N, Homuth G, Hammer E, Völker U. Associations of circulating plasma microRNAs with age, body mass index and sex in a population-based study. BMC Med Genomics 2015;8(1):1-9.
18. Lewis CA, Layne TR, Seashols-Williams SJ. Detection of microRNAs in DNA extractions for forensic biological source identification. J Forensic Sci 2019;64(6):1823-30.
19. Centers for Disease Control and Prevention. All About Adult BMI – 2021. https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html (accessed February 26, 2021).
20. Chen X, Liang H, Guan D, Wang C, Hu X, Cui L, Chen S, Zhang C, Zhang J, Zen K, Zhang CY. A combination of Let-7d, Let-7g and Let-7i serves as a stable reference for normalization of serum microRNAs. PloS One 2013;8(11):e79652.
21. RStudio Team. RStudio: Integrated Development Environment for R [Internet]. Boston, MA; 2020. Available from: http://www.rstudio.com/
22. Rothman KJ. BMI-related errors in the measurement of obesity. Int J Obes 2008;32(3):S56-9.

23. Schmidt WM, Spiel AO, Jilma B, Wolzt M, Müller M. In vivo profile of the human leukocyte microRNA response to endotoxemia. Biochem Biophys Res Commun. 2009;380(3):437-41.

24. Chen SY, Wang Y, Telen MJ, Chi JT. The genomic analysis of erythrocyte microRNA expression in sickle cell diseases. PloS One. 2008;3(6):e2360.

25. Purdy JC, Shatzel JJ. The hematologic consequences of obesity. Eur J Haematol 2021;106(3):306-19.

26. Vu L, Ragupathy V, Kulkarni S, Atreya C. Analysis of Argonaute 2–microRNA complexes in ex vivo stored red blood cells. Transfusion. 2017 Dec;57(12):2995-3000.

27. Kim YJ, Hwang SH, Lee SY, Shin KK, Cho HH, Bae YC, Jung JS. miR-486-5p induces replicative senescence of human adipose tissue-derived mesenchymal stem cells and its expression is controlled by high glucose. Stem Cells Dev 2012;21(10):1749-60.

28. Hirsch J, Batchelor B. Adipose tissue cellularity in human obesity. Clin Endocrinol Metab 1976;5(2):299-311.

29. Raitoharju E, Seppälä I, Lyytikäinen LP, Viikari J, Ala-Korpela M, Soininen P, Kangas AJ, Waldenberger M, Klopp N, Illig T, Leiviskä J. Blood hsa-miR-122-5p and hsa-miR-885-5p levels associate with fatty liver and related lipoprotein metabolism—The Young Finns Study. Sci Rep 2016;6(1):1-3.

30. Fan R, Wang J, Du J. Association between body mass index and fatty liver risk: A dose-response analysis. Sci Rep 2018;8(1):1-7.

31. Giroud M, Karbiener M, Pisani DF, Ghandour RA, Beranger GE, Niemi T, Taittonen M, Nuutila P, Virtanen KA, Langin D, Scheideler M. Let-7i-5p represses brite adipocyte function in mice and humans. Sci Rep 2016;6(1):1-1.

32. Chen YQ, Wang XX, Yao XM, Zhang DL, Yang XF, Tian SF, Wang NS. Abated microRNA-195 expression protected mesangial cells from apoptosis in early diabetic renal injury in mice. J Nephrol 2012;25(4):566-76.
33. Wen D, Qiao P, Wang L. Circulating microRNA-223 as a potential biomarker for obesity. Obes Res Clin Pract 2015;9(4):398-404.
34. Iannone F, Montesanto A, Cione E, Crocco P, Caroleo MC, Dato S, Rose G, Passarino G. Expression patterns of muscle-specific miR-133b and miR-206 correlate with nutritional status and sarcopenia. Nutrients 2020;12(2):297.
35. Ortega FJ, Moreno-Navarrete JM, Pardo G, Sabater M, Hummel M, Ferrer A, Rodriguez-Hermosa JI, Ruiz B, Ricart W, Peral B, Fernández-Real JM. MiRNA expression profile of human subcutaneous adipose and during adipocyte differentiation. PloS One 2010;5(2):e9022.
36. Ostrowska L, Gornowicz A, Pietraszewska B, Bielawski K, Bielawska A. Which salivary components can differentiate metabolic obesity? Plos One 2020;15(6):e0235358.
37. de Heredia FP, Gómez-Martínez S, Marcos A. Obesity, inflammation and the immune system. Proc Nut Soc 2012;71(2):332-8.

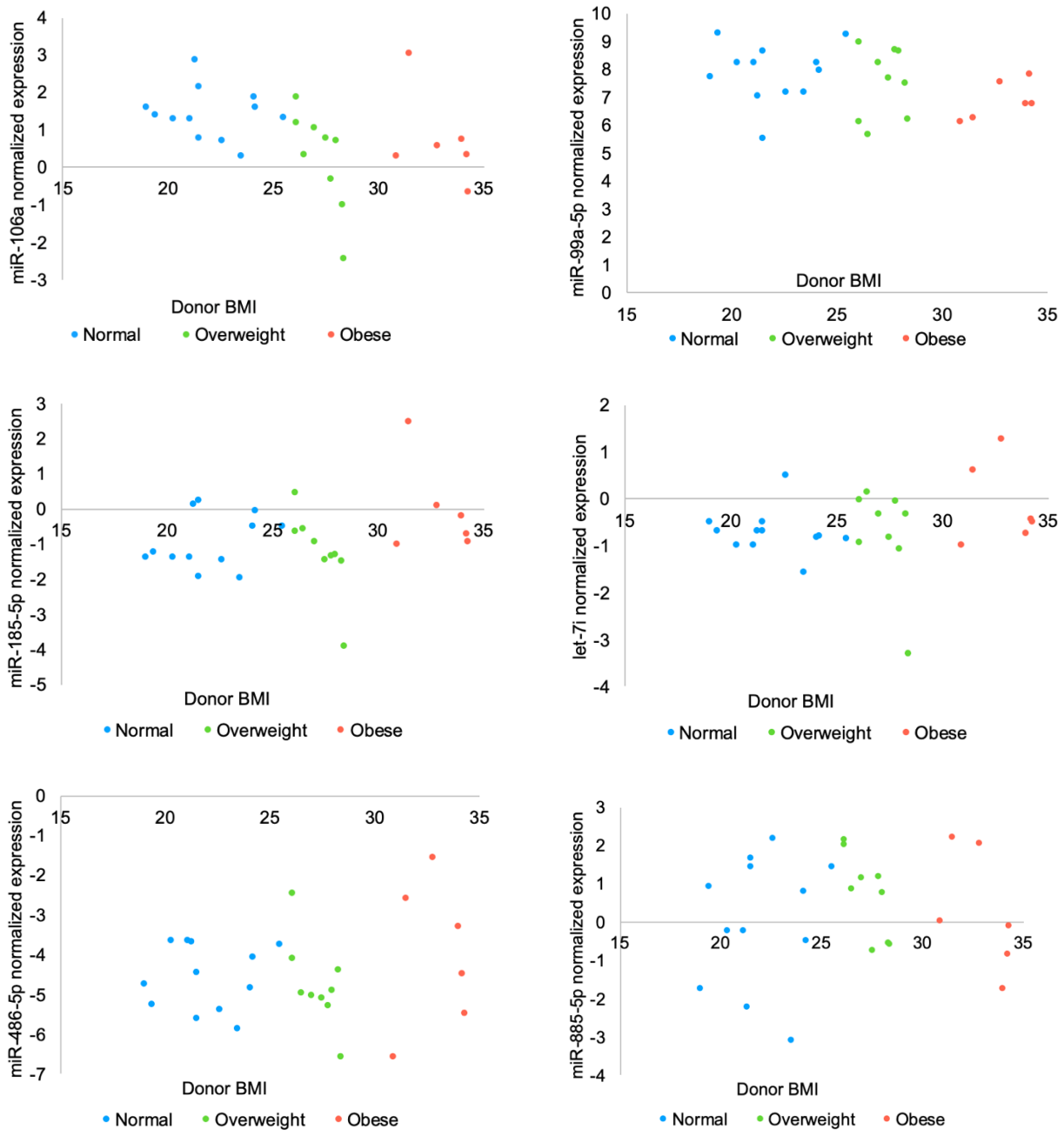


Figure 1. miRNA expression from preliminary study using blood serum and an RNA isolation method and samples with known BMI ranging from normal (BMI = 18.5-24.9), overweight (BMI = 25-29.9), to obese (BMI ≥ 30). Expression normalized to let-7g.

Table 1. Selected samples

Sample	Sex	Race	Weight (lbs)	Height (in)	BMI
7046	Male	Caucasian	115	69	17.0
5117	Female	Asian	106	65	17.6
8018	Female	Asian	105	64	18.0
1222	Female	African American	123	69	18.2
8021	Female	Caucasian	130	65	21.6
8026	Female	African American	119	62	21.8
1087	Female	Asian	125	63	22.1
8004	Female	Hispanic	135	65	22.5
5137	Female	Caucasian	145	67	22.7
8016	Female	Caucasian	150	68	22.8
8020	Female	Caucasian	146	66	23.6
1213	Male	Asian	160	69	23.6
1088	Female	African American	140	64	24.0
8008	Male	Caucasian	175	68	26.6
7042	Male	African American	175	66	28.2
1090	Female	Caucasian	162	63.5	28.2
1081	Female	Caucasian	170	65	28.3
8031	Female	Asian	150	61	28.3
7505	Male	Asian	198	70	28.4
1089	Female	Hispanic	180	66	29.0
7504	Female	Asian	180	65	30.0
7507	Female	African American	180	65	30.0
1220	Male	Caucasian/Pacific Islander	225	71	31.4
8015	Female	Chinese	125	52	32.5
1085	Female	Caucasian	210	66	33.9

Blood samples selected from the IRB-approved registry and used for testing. BMI is colored based on category with green representing underweight individuals (BMI < 18.5) (4), yellow representing normal weight individuals (BMI = 18.5-24.9) (9), orange representing overweight individuals (BMI = 25-29.9) (7), and red representing obese individuals (BMI ≥ 30) (5).

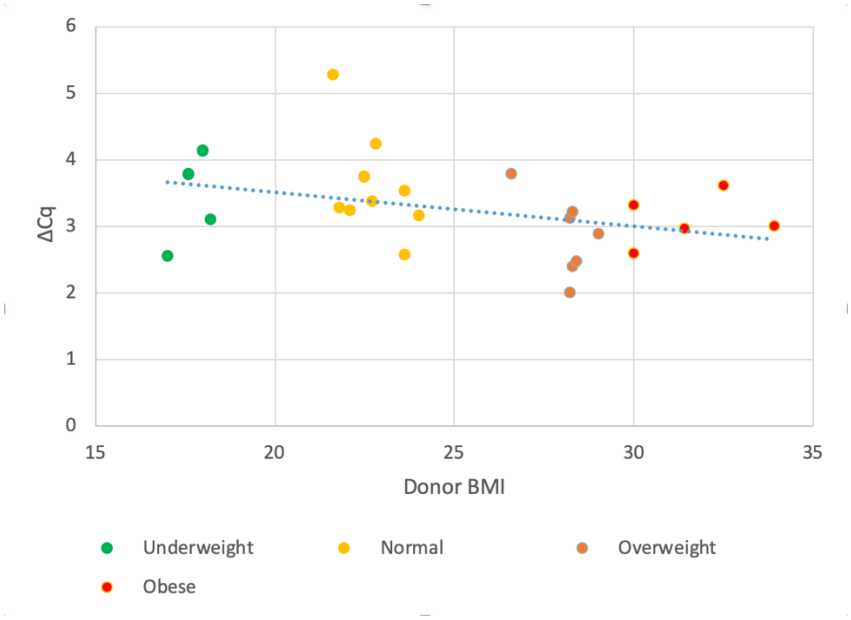
Table 2. Candidate miRNAs

miRNA ID	Accession #	Human miRNA Sequence	Primer Sequence
hsa-miR-20a-5p*	MIMAT0000075	UAAAGUGCUUAUAGUGCAGGUAG	CGCTAAAGTGCTTATAGTGCAGGT
hsa-miR-93-5p*	MIMAT0000093	CAAAGUGCUGUUCGUGCAGGUAG	AAAGTGCTGTTCGTGCAGGT
hsa-miR-99a-5p**	MIMAT0000097	AACCCGUAGAUCCGAUCUUGUG	CGCCAACCCGTAGATCC
hsa-miR-106a-5p	MIMAT0000103	AAAAGUGCUUACAGUGCAGGUAG	CGCCAAAAGTGCTTACAGTGC
hsa-miR-143-3p*	MIMAT0000435	UGAGAUGAAGCACUGUAGCUC	TGAGATGAAGCACTGTAGCTCAAA
hsa-miR-145-5p	MIMAT0000437	GUCCAGUUUUCCCAGGAAUCCCU	CAGTTTTCCCAGGAATCCCTAA
hsa-miR-148a-3p*	MIMAT0000243	UCAGUGCACUACAGAACUUUGU	CGCTCAGTGCACTACAGAACTTT
hsa-miR-185-5p	MIMAT0000455	UGGAGAGAAAGGCAGUUCCUGA	TGGAGAGAAAGGCAGTTCCTG
hsa-miR-194-5p	MIMAT0000460	UGUAACAGCAACUCCAUGUGGA	ACAGCAACTCCATGTGGAAAA
hsa-miR-215-5p	MIMAT0000272	AUGACCUAUGAAUUGACAGAC	CGCATGACCTATGAATTGACAGAC
hsa-miR-365a-3p*	MIMAT0000710	UAAUGCCCCUAAAAAUCCUUAU	CGTAATGCCCCTAAAAATCCTT
hsa-miR-486-5p	MIMAT0002177	UCCUGUACUGAGCUGCCCCGAG	GAGCTGCCCCGAGAAAAA
hsa-miR-505-3p	MIMAT0002876	CGUCAACACUUGCUGGUUUCCU	CGTCAACACTTGCTGGTTTCC
hsa-miR-885-5p	MIMAT0004947	UCCAUUACACUACCCUGCCUCU	TCCATTACACTACCCTGCCTCT
hsa-let-7i-5p	MIMAT0000415	UGAGGUAGUAGUUUGUGCUGUU	CGTTCTGAGGTAGTAGTTTGTGCT

*excluded from combined (Cq and ΔCq) support-vector machine analysis due to missingness from some samples

**excluded from all statistical analyses due to lack of expression in all but two samples

A.



B.

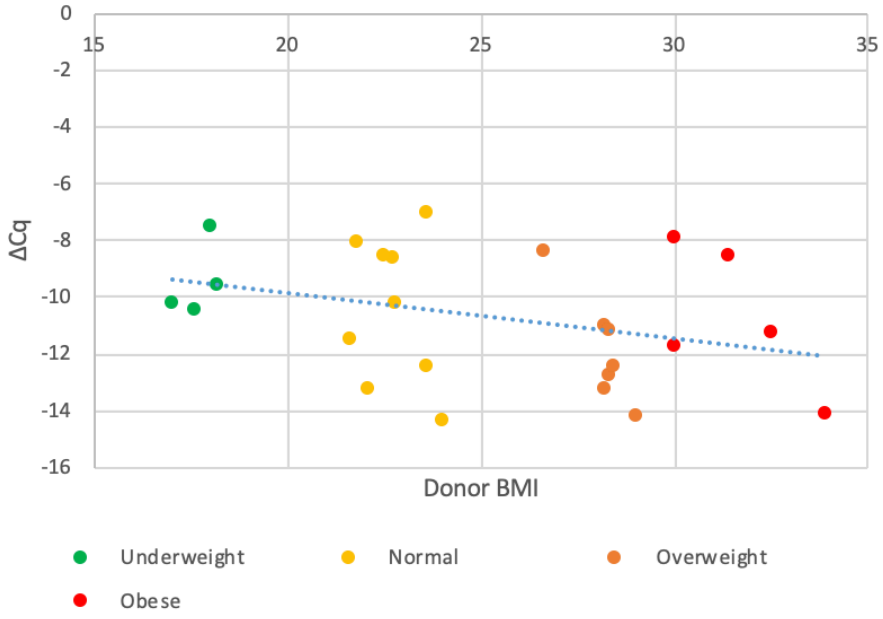


Figure 2. Observed relationships between ΔCq and BMI
Calculated ΔCq using let-7g as the normalizer plotted against the known donor BMI for (A) miR-486-5p and (B) miR-885-5p. Samples are colored according to BMI classification with underweight (green), normal (yellow), overweight (orange), and obese (red). Trendline added to show potential negative association between miRNA expression and BMI.

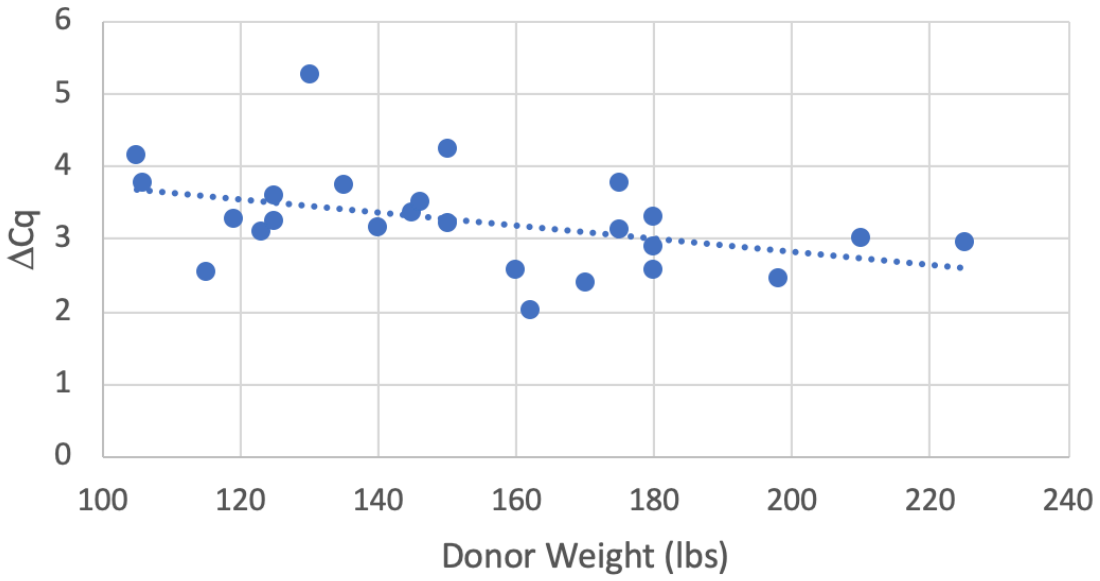
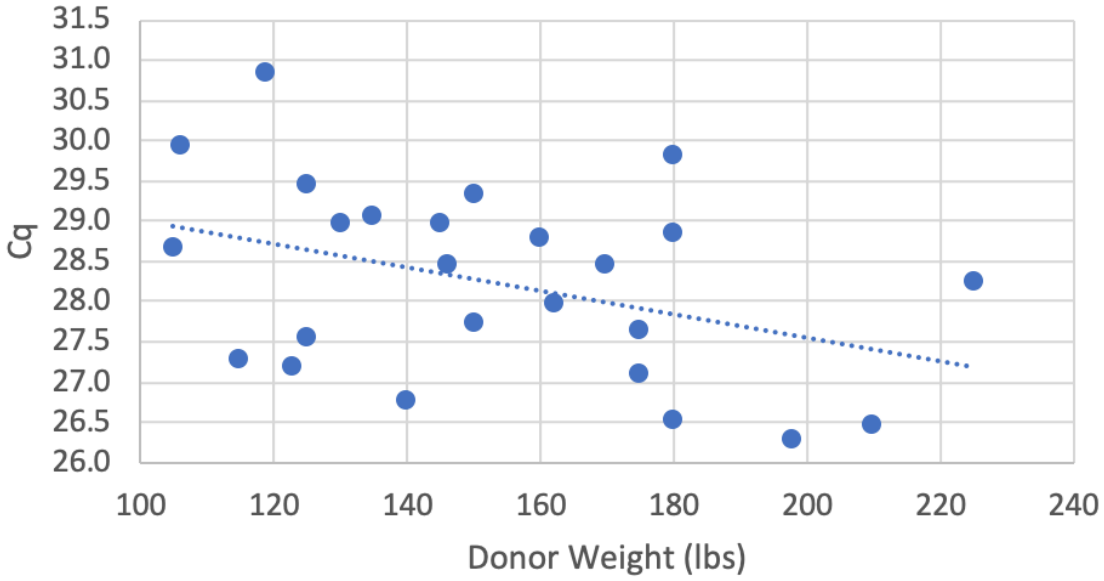


Figure 3. Observed relationship between ΔCq and weight
Calculated ΔCq using let-7g as the normalizer plotted against the known donor weight for miR-486-5p. Trendline added to show potential negative association between miRNA expression and donor weight.

A.



B.

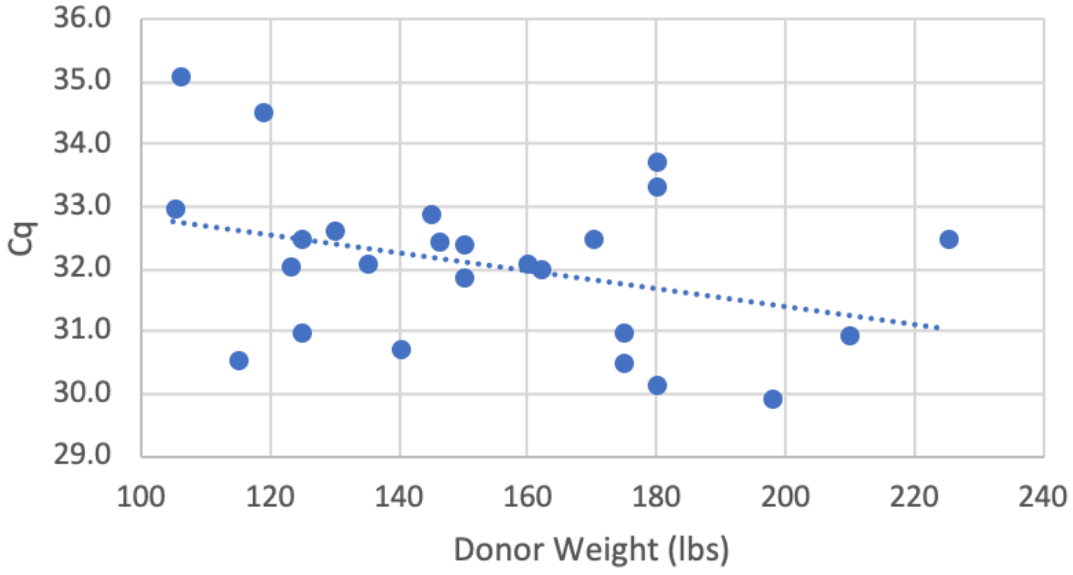


Figure 4. Observed relationship between Cq and weight
Observed Cq plotted against the known donor weight for (A) let-7i-5p and (B) miR-194-5p. Trendline added to show potential negative association between miRNA expression and weight.

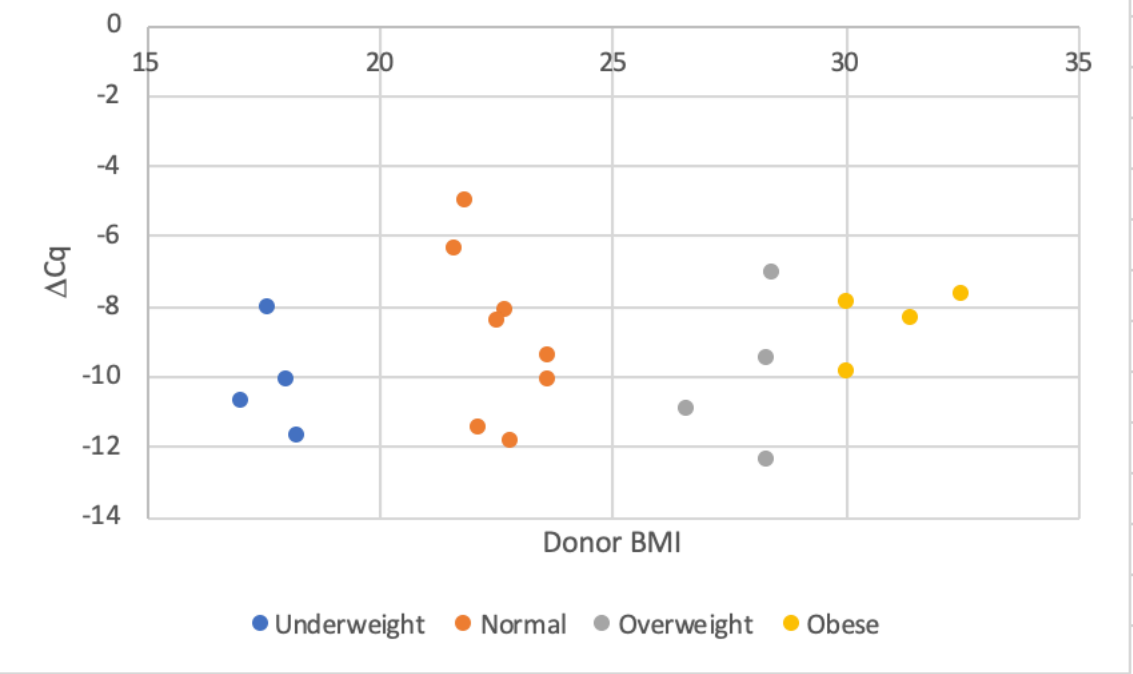


Figure 5. Example miRNA of no interest
Example of miRNA marker (miR-143-3p) showing no trend of interest when calculated ΔCq was plotted against the known donor BMI. Samples are colored according to BMI classification with underweight (green), normal (yellow), overweight (orange), and obese (red).

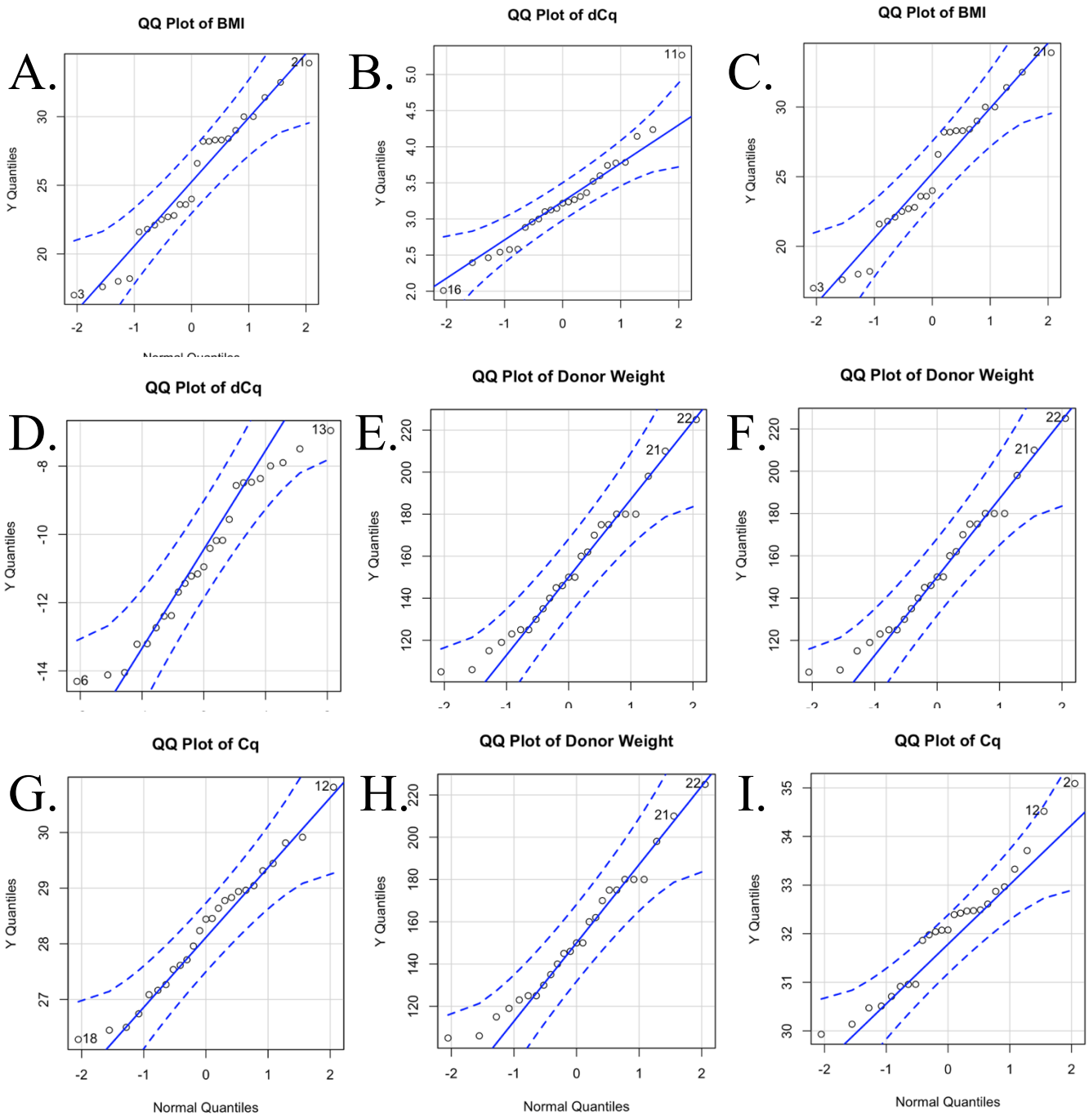


Figure 6. Q-Q plots for association analyses
Quantile-quantile (Q-Q) plots showing linearity and homoscedasticity for (A) BMI and (B) ΔCq observations from miR-486-5p dataset, (C) BMI and (D) ΔCq from the miR-885-5p dataset, (E) donor weight from the miR-486-5p dataset, (F) donor weight and (G) Cq from the let-7i-5p dataset, and (H) donor weight and (I) Cq from the miR-194-5p dataset.

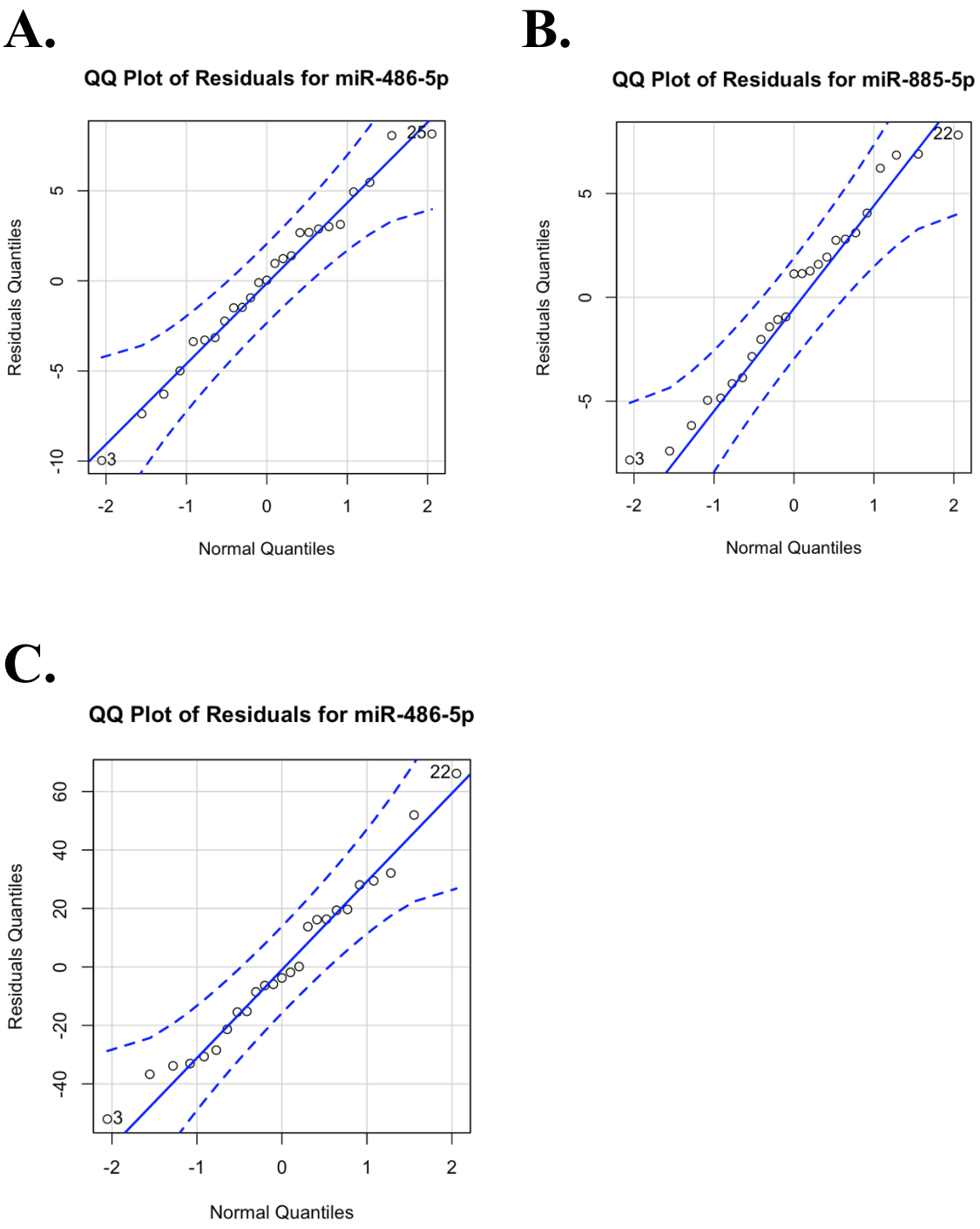
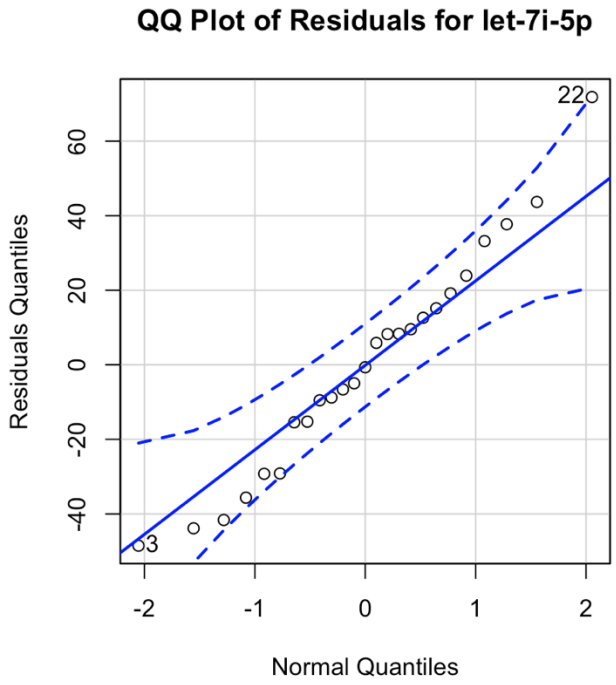


Figure 7. Q-Q plots of residuals for associations with ΔCq
Quantile-quantile (Q-Q) plot of residuals from simple linear regression analysis between (A) calculated ΔCq values and BMI for miR-486-5p, (B) miR-885-5p. Q-Q plot of residuals between (C) ΔCq values and donor weight for miR-486-5p also shown.

A.



B.

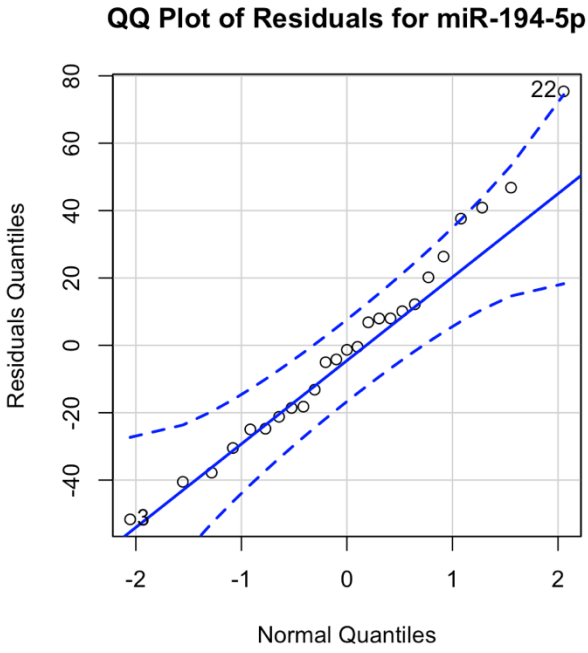


Figure 8. Q-Q plot of residuals for associations with Cq
Quantile-quantile plot of residuals from simple linear regression analysis between observed Cq values and donor weight of (A) let-7i-5p expression and (B) miR-194-5p.

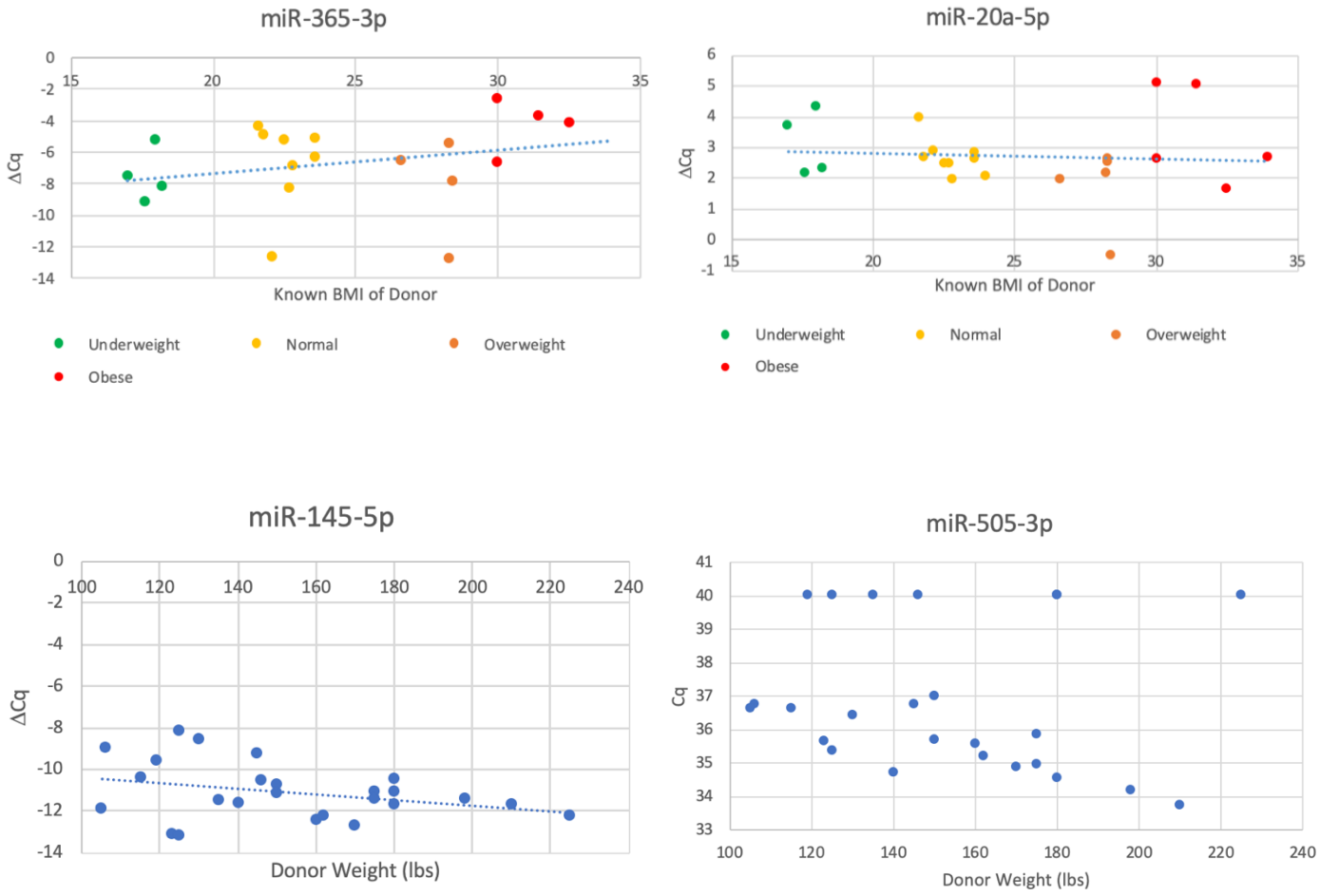
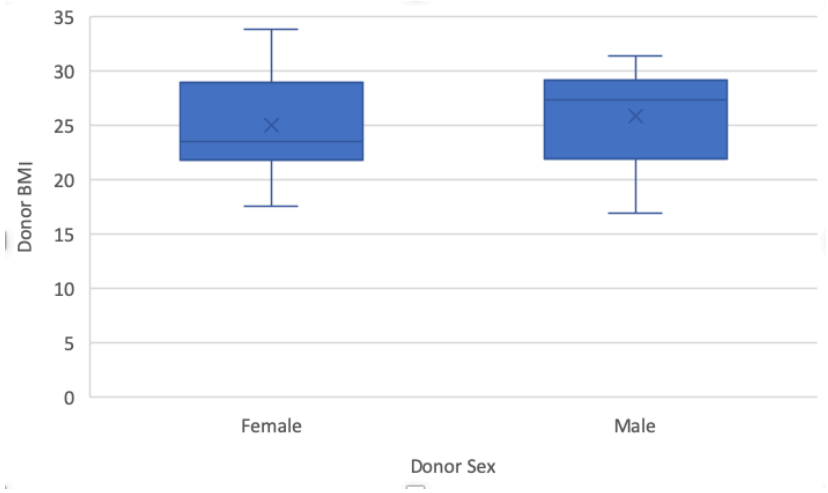


Figure 9. Observations from preliminary study
Trends observed in preliminary visualization of data found to not be statistically significant for miR-365-3p, miR-20a-5p, miR-145-5p, and miR-505-3p. Samples in plotted against BMI are colored according to BMI classification with underweight (green), normal (yellow), overweight (orange), and obese (red). Trendlines added where appropriate to show potential trends.

A.



B.

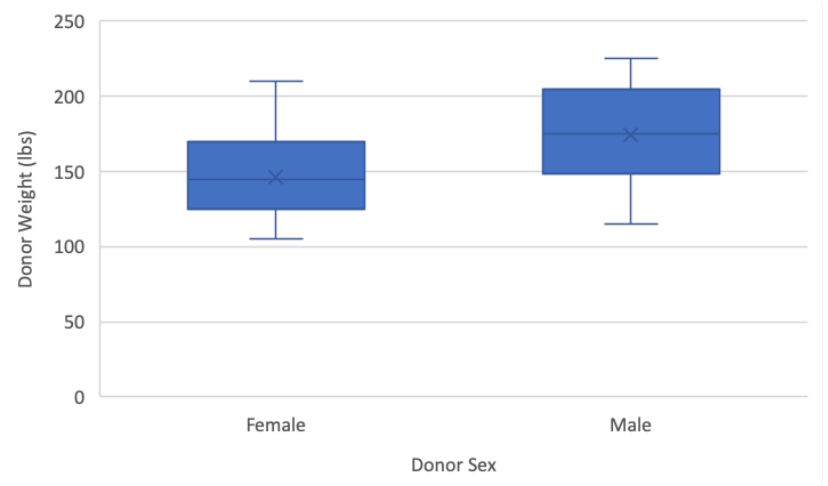


Figure 10. Observed differences between sexes for BMI and weight
Boxplots showing observed differences between (A) reported donor sex and BMI, as well as (B) donor weight. Differences in distribution of female vs male donor weight due to more females represented in sample set. 19 females and six males were included in the sample set.

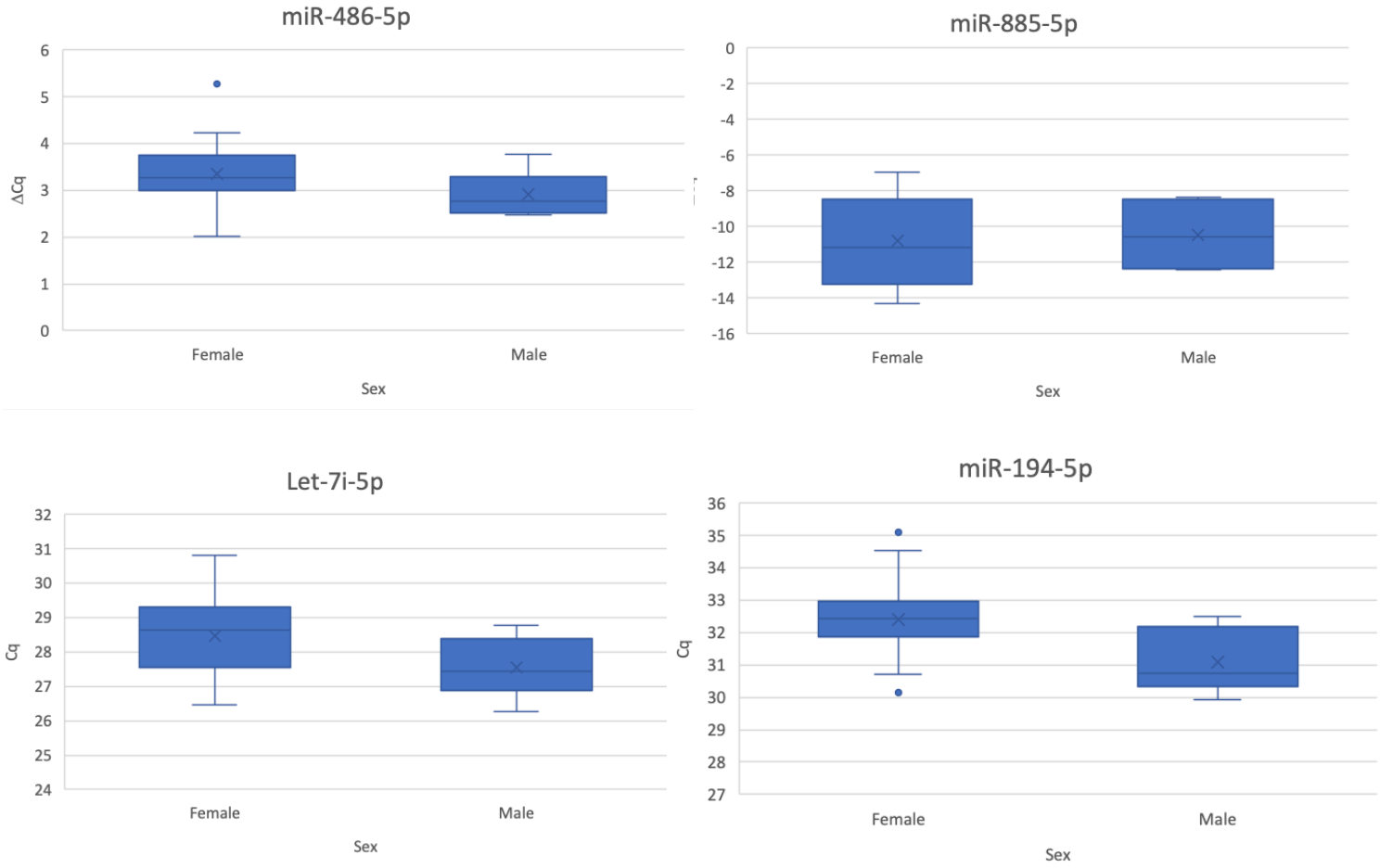
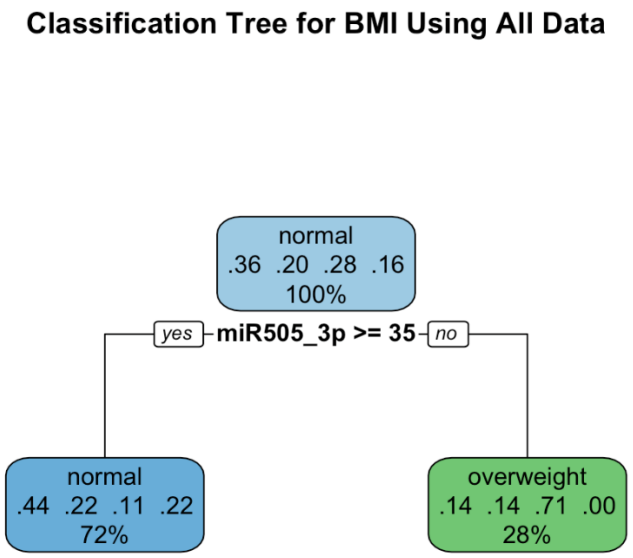


Figure 11. Observed differences between sexes for analyzed miRNAs
Boxplots showing observed distribution of either ΔCq or Cq of miRNA markers with statistically significant association to either BMI or donor weight with reported sex of donor. No differences identified between sexes were noted as unexpected for any miRNA markers.19 females and six males were included in the sample set.

A.



B.

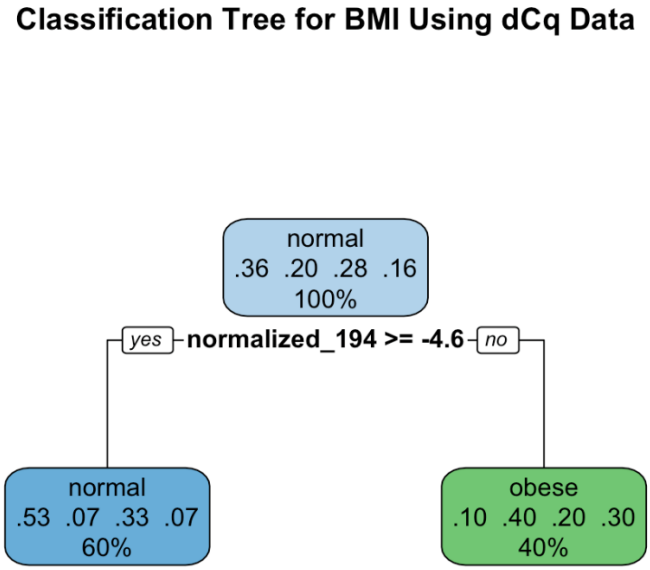


Figure 12. Graphical representation of CART predictions
Visualization of CART analyses using both raw Cq and ΔCq from all markers (A), as well as just ΔCq data (B).

Table 3. BMI predictions from CART analyses

A.

Actual	Prediction			
	Underweight	Normal	Overweight	Obese
Underweight	0	4	0	0
Normal	0	8	1	0
Overweight	0	2	5	0
Obese	0	4	1	0

B.

Actual	Prediction			
	Underweight	Normal	Overweight	Obese
Underweight	0	1	0	3
Normal	0	8	0	1
Overweight	0	5	0	2
Obese	0	1	0	4

Categorical BMI prediction of samples compared to known BMI categorization using CART analysis for (A) all data (52% correct) and (B) just ΔCq data (48% correct).

Table 4. BMI predictions from SVM analyses

A.

Actual	Prediction			
	Underweight	Normal	Overweight	Obese
Underweight	4	0	0	0
Normal	0	9	0	0
Overweight	0	0	7	0
Obese	0	0	0	5

B.

Actual	Prediction			
	Underweight	Normal	Overweight	Obese
Underweight	4	0	0	0
Normal	0	6	0	0
Overweight	1	0	3	0
Obese	0	0	0	4

Categorical BMI prediction using the support vector machine model using (A) combined Cq and ΔCq data (100% correct) and (B) just raw Cq data (94% correct). SVM analysis using combined data had 22 support vectors and the SVM analysis had 18 support vectors. Due to sample missingness, miRNA markers with any missing values were eliminated from the combined analysis. Markers used for the combined analysis were: miR-106a-5p, miR-145-5p, miR-185-5p, miR-194-5p, miR-215-5p, miR-486-5p, miR-505-3p, miR-885-5p, let-7i-5p. All markers were used for analysis consisting only of Cq values, which contributed to samples being excluded and the N=18 as shown in B.

Supplemental Table 1. List of R code used for each analysis in RStudio for this study

Function	R Code
Quantile-Quantile Plot	qqPlot()
Simple/Multiple Linear Regression	lm()
Classification and Regression Tree	rpart() rpart.plot()
Support Vector Machine	svm()

Vita

Jennifer Shealy is originally from Prosperity, South Carolina and graduated from Mid-Carolina High School in 2015. She attended Clemson University where she majored in genetics and received a minor in both biological sciences and psychology. While at Clemson, she worked in a biochemistry lab with the purpose of identifying drug targets for a kinetoplastid parasite. She presented her research titled "An Investigation of PEX 13.1 and PEX 13.2 Proteins in Trypanosoma brucei" at the Summer CI and Fall Kick-Off event at Clemson University. After completing her Bachelor of Science in 2019, she enrolled at Virginia Commonwealth University, joining the Department of Forensic Science graduate program in the forensic biology concentration. Starting in the fall of 2020, she began working in the Williams lab working on research involving miRNA targets for phenotyping purposes. During her time at VCU, she taught multiple undergraduate biology lab courses, introducing basic biology concepts to both science majors and students majoring in non-science courses.