Circulating microRNAs and Both Association With Methacholine PC20 and Prediction of Asthma Exacerbation in the Childhood Asthma Management Program (CAMP) Cohort

Joshua Seth Davis, M.D.
Master of Medical Sciences in Biomedical Informatics Thesis

Circulating microRNAs and both association with methacholine PC_{20} and prediction of asthma exacerbation in the Childhood Asthma Management Program (CAMP) Cohort

Dedication

This thesis is dedicated to my parents – Ronald P. Davis and Linda H. Davis. Without their support and love, I would not be the person who I am today! Thank you both for everything that you have done for me.

Acknowledgments

Drs. Sun and Lu contributed the functional validation data in Chapter 3.
The below individuals also helped contribute to this work:

Alvin T. Kho, Ph.D[1,2]
Maoyun Sun, Ph.D[3]
Michael McGeachie, Ph.D[1]
Kip G. Moore, BA[1]
Jody M. Sylvia, MS[1]
Quan Lu, Ph.D[3,*]
Scott T. Weiss, M.D., MPH[1,*]
Kelan G. Tantisira, M.D., MPH[1,*,#]

[1] Department of Medicine, Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[2] Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA

[3] Department of Environmental Health and Department of Genetics and Complex Diseases, Program in Molecular and Integrative Physiological Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[*] Thesis Committee members
[#] Thesis Committee Chair

Funding

This work is supported by the NIH grants R01 HL127332, R01 HL129935, U01 HL65899, and R01 NR013391. AK is supported via K25 HL091124. J.D. is supported by T32 HL007427. Q.L. is supported in part by funding from the US National Institute of

Health (NIH, R01HL114769) and by the Scholar Award from the American Asthma Foundation. M.Y.S. is supported by an NIH training grant (T32HL007118).

Declaration of Interests

None

Legal issues

Any questions regarding copyrights or legal matters should be referred to the Brigham and Women's Hospital legal department.

Foreword

Note: The below foreword represents my own thoughts and ideas and not those necessarily of Brigham and Women's Hospital or Harvard Medical School.

I thought placing an "easter egg" in my thesis was a great idea to add some non-scientific knowledge to the Countway Library of Medicine and record an indelible "thank you" to all those in the Boston area that have participated in not only my Master's education but also my medical education over these past seven years.

Education comes in many forms, and one influential and non-academic source in my generation was the Legend of Zelda video games. In the halcyon days of the Reagan era, Nintendo released in 1986 the first game, "The Legend of Zelda." The series has received wide acclaim and recently celebrated its 31st anniversary in 2017 with the release of "Breath of the Wild." In the first game of the series, the protagonist Link walks into a cave in which he finds a bearded elderly man standing between two flames. The man states, "IT'S DANGEROUS TO GO ALONE! TAKE THIS." It was not only a powerful message but also the beginning to what would become a key series for Nintendo. Link would continue to explore various aspects of the open world finding further hidden surprises, and the player clearly learned the value of going off the beaten path and exploring. Link would often go into villages, speak with various inhabitants, and obtain advice or tools for the next step of his journey. His quest was ultimately to save Princess Zelda after evil had once again appeared in the land. One exception to this premise was Majora's Mask, which was a darker game and emphasized a personal quest. Shigeru Miyamoto, the creator of the series, would later reveal that this exploration of the natural world was modeled after his TV-less childhood and incorporated this into the Zelda series. Although other games offered exploration, there seemed to be something special about this particular series.

Many non-Japanese partaking in the Legend of Zelda may not appreciate the extent of Japanese mythology incorporated into the series, but these elements are likely what makes the series so intriguing. While Joseph Campbell was one of the foremost scholars of mythology, Fanny Mayer was able to compile a delightful collection of Japanese tales respecting the rich oral tradition of these stories passed from generation to generation. Mayer [1] describes a familiar landscape to the Zelda games in her book, *Ancient Tales in Modern Japan*: "Men get ahead in life by the work of their hands. The man that breaks sod for his garden patch in the hills is often rewarded by the unexpected discovery of treasure… there is virtue in the treasure that brings prosperity. And prosperity does not make him a "big shot." He and his wife will live without worry about food and clothes, but frequently their blessing is a lovely bride for an unpromising son… The world of the Japanese folk tale is its land, its mountains, waterways, plants, animals, and the changes of its seasons. It is a world in which man is not dominant. It is shared by all life, visible and invisible. Deities are close at hand to hear petitions… Spirits of the dead remain close by in the mountains to watch over the needs of their families. Mountains are not just a backdrop. Trees in them furnish firewood and nuts, and other plants in them provide food to be gathered. Paths through the mountains are fraught with danger. Water in the tales is found in mountain springs or pools and in streams in little valleys. Guardian spirits in the water may be either benevolent or malevolent. The seashore or the sea also opens adventure to characters." These concepts serve as a blueprint for the Zelda games. The real world itself is a complex place, and there are moments or experiences when the world pulls you into its mysteries and intricacies making exploration and the associated struggle worthwhile. Oftentimes in Link's world, the rewards can be practical (e.g. a boomerang) or bring knowledge (e.g. a song passed on to the player). Sometimes these rewards are less practical and may simply be an opportunity to see or hear something that adds perspective and makes the world more complete. Another important myth is the "Warashibe Choja" or the Straw Millionaire, which is incorporated into the series with trading or complex exchanges between characters to achieve an outcome. Notably in Majora's Mask, completed tasks resulted in a different ending depending on what the player achieved. If Link helped many people and completed side quests, then the ending was more fruitful for the player. Overall, these concepts make the game special.

How does this all exactly related to education or this thesis? Although many of us are not out to save a princess, the Zelda series and the underlying Japanese mythology offer a roadmap for learning, education, and life and were formative in my development. At least from my observations, the issue with the world today is that there is a growing mass of the population with ennui and a dearth of optimism, wonder, and curiosity. I think that people have trouble appreciating the mystery of existence and the world around them for whatever reason. Away from a computer or smartphone, there is the world waiting to be discovered, and one may even encounter the unexpected or meet someone that may provide further insight. Like Ferris Bueller concluded at the end of *Ferris Bueller's Day Off* (1986), "Life moves pretty fast. If you don't stop and look around once in a while, you could miss it." Even with bioinformatics and the increasing amount of data available for analyses, nobody alone has all the answers and never will. There is no master algorithm, all-knowing person, or computer due the quantum mechanical

properties of matter and resultant inherent uncertainty. We must continue to learn from each other, and despite recent geopolitical tensions in the world, we are all in this together.

I have been on a journey myself encompassing broad training in medicine and biomedical research during my time in Boston. I am fortunate to have had this opportunity, but I did not complete this journey myself. Many of you probably never realized that like the old man that offered Link something for his journey, you have offered me something to take – it might have been advice, knowledge, skills, support, humor, coffee or something else. For these things, I am truly grateful and hope that I was also able to offer something in return. Thank you for all that I have learned from you - that has been my richest reward during my time in Boston.

Take care,

Joshua Seth Davis, M.D.

Reference:

1. Mayer, Fanny H. *Ancient tales in modern Japan: anthology of Japanese folk tales*. Indiana University Press, 1985.

Abstract

Background: Circulating microRNAs have shown promise both as a non-invasive biomarker and a predictor of disease activity. Prior asthma studies with clinical, biochemical, or genomic tests have failed to demonstrate excellent prediction of asthma exacerbation. This thesis hypothesizes that circulating miRNA would reveal: 1) candidate biomarkers related to airway hyperresponsiveness (AHR) and provide biologic insights into asthma epigenetic influences, 2) a panel of circulating miRNA in a pediatric asthma cohort or a combined microRNA-clinical asthma exacerbation score that may have superior predictive capability compared to the clinical asthma exacerbation score alone, and 3) the power of machine learning techniques such as backpropagation neural networks and XGBoost that may provide both improved prediction of exacerbations or perhaps yield insight into complex non-linear behavior.

Methods: Serum samples were obtained at randomization for 160 children in the Childhood Asthma Management Program and were profiled using the TaqMan miRNA array set containing 754 miRNAs. For Aims 2 and 3, only 153 subjects had complete information on steroid bursts in the first year after randomization.

Aim 1: The association of the isolated miRNA with methacholine PC_{20} was assessed. Network and pathway analyses were performed. Functional validation of two significant miRNAs was performed in human airway smooth muscle cells (HASMs).

Aim 2: Dichotomized data for asthma exacerbation from the first year after randomization to the inhaled corticosteroid arm were used for binary logistic regression

with microRNA cycle threshold and clinical exacerbation score. Ontology and pathway analyses were performed for significant miRNAs

Aim 3: Dichotomized data for asthma exacerbation as in Aim 2 was used for both backpropagation and resilient backpropagation neural networks and XGBoost with miRNA features selected by the Kruskal Test for binary and multiclass classification. Randomized hyperparameter optimization was used for all 3 algorithms. The neural network metric for classification was mean misclassification error. The XGBoost metric for classification was AUROC.

Results:

Aim 1: Of 155 well-detected circulating miRNAs, eight were significantly associated with PC_{20} with the strongest association with miR-296-5p. Pathway analysis revealed miR-16-5p as a network hub, and involvement of multiple miRNAs interacting with genes in the FoxO and Hippo signaling pathways by KEGG analysis. Functional validation of two miRNA in HASM showed effects on cell growth and diameter.

Aim 2: Of the 125 well-detected circulating miRNA, 12 had significant odd ratios for exacerbation with the most significant being miR-206. Each doubling of expression of the 12 miRNA resulted in between a 25-67% increase in risk of exacerbation. Stepwise logistic regression resulted in a three miRNA model that, when combined with the clinical score, demonstrated an AUROC of 0.81, which was superior to either the clinical model alone (AUROC 0.67) or miRNA model (AUROC 0.71). The three microRNAs also had biological relevance with involvement in NK-kB signaling and inactivation of Gsk3 by AKT pathways.

Aim 3: For the neural network analysis, the backpropagation algorithm with all miRNA features had an improved mean misclassification error (0.248) compared to the resilient backprogation algorithm with all features and the top 10 features. The XGBoost model had an AUROC of 0.69, which was similar to the AUROC of the 3 miRNA panel found in the logistic regression analysis (Aim 2).

Conclusion:

Aim 1: Reduced circulatory miRNA expression at baseline is associated with an increase in PC_{20}. These miRNA provide biologic insights into, and may serve as biomarkers of, asthma severity. miR-16-5p and -30d-5p regulate airway smooth muscle phenotypes critically involved in asthma pathogenesis, supporting a mechanistic link to these findings. Functional ASM phenotypes may be directly relevant to AHR.

Aim 2: Circulating microRNAs combined with a clinical model of asthma exacerbations demonstrate prediction of exacerbation status in subjects taking inhaled corticosteroids with use of logistic regression.

Aim 3: These preliminary analyses with machine learning algorithms including neural network and XGBoost did not result in improved prediction of asthma exacerbation compared to logistic regression. Neural networks were computationally complex and did not result in improved accuracy compared to the no information rate. XGBoost and logistic regression results used different feature selection (non-parametric vs. parametric) and data pre-processing steps (dense matrix input vs. median imputation of missing values) yielding similar predictive capability for asthma exacerbation with circulating miRNA. While further analyses are needed, these preliminary results support robust prediction with circulating miRNA especially given miR-206 was a key feature among models despite different feature selection and analytic techniques.

Chapters

Chapter 1 – Introduction
Chapter 2 – Materials and Methods
Chapter 3 – Results and Discussion for association study with circulating microRNAs and methacholine PC_{20}
Chapter 4 – Results and Discussion for prediction of asthma exacerbation using logistic regression with panel of circulating microRNAs compared to an asthma exacerbation clinical score
Chapter 5 – Results and Discussion for prediction of asthma exacerbation with panel of circulating microRNAs with additional machine learning techniques
Chapter 6 – Conclusion and Future Directions
Appendices

Chapter 1 – Introduction and Aims

A. Introduction

Asthma is a chronic inflammatory respiratory disease that affects greater than 300 million people worldwide [1]. Asthma is characterized by airway obstruction due to a combination of smooth muscle hyperresponsiveness and inflammation [2]. Both asthma therapies and hospitalizations/doctor visits generate significant healthcare utilization with an estimated $ 62.8 billion dollars spent on diagnosis and management of asthma in the United States in 2009 [3,4]. It remains challenging to generate risk assessment, predict prognosis, and determine optimal treatment response in asthmatics. Further elucidation of mechanisms associated with airways hyperresponsiveness (AHR) could foster further biological knowledge especially in the evolving area of epigenetic influences. Prediction of important clinical events such as asthma exacerbation could also result in aggressive preventive care, improved clinical outcomes, and reduced healthcare costs. Prior studies including biochemical markers [5], clinical history such as prior severe exacerbations [6], and genomic data such as SNPs [7] have yielded variable prediction capability.

An important area of epigenetic investigation in respiratory disease is the study of microRNAs (miRNA). Circulating microRNAs (miRNAs) are promising biomarkers for human diseases [8] and may be helpful in a variety of clinical scenarios from risk assessment to monitoring response to treatment [9]. miRNA characteristics and function have been well described in the literature [10]. In brief, miRNAs are a class of small RNAs that inhibit gene expression by binding to the 3'-untranslated region (UTR) of messenger RNAs to degrade or suppress the translation of the mRNA. Given the

availability of miRNA mimics and antagonists, these small RNAs have been proposed as therapeutic targets. Circulating miRNAs are highly stable in the serum [11]. miRNA plasma biomarkers have been proposed for neurological conditions [12], cancer detection/prognosis [13], cardiovascular disease [14], and other conditions including an emerging role in respiratory diseases [15]. Translational methods have been applied in order to generate screening tests [16]. For instance, a recent study demonstrated promise with use of a subset of circulating miRNA expressed uniquely between asthma and allergic rhinitis patients [17]. miRNAs may also be associated with other important traits such as AHR and have predictive power for important clinical events such as asthma exacerbations.

B. Hypothesis

The hypothesis is that circulating miRNA in a well-characterized cohort are associated with asthma severity as determined by methacholine PC_{20} and also can serve as a non-invasive biomarker to predict significant clinical events like asthma exacerbation both with traditional methods such as logistic regression and machine learning techniques. Resources for this thesis include both extensive clinical data from the Childhood Asthma Management Program (CAMP) cohort [18] and miRNA profiling of the serum in a subset of 160 Caucasian subjects in CAMP.

C. Specific Aims

Aim 1: I hypothesize that the association of circulating miRNA with a measure of airways hyperresponsiveness (methacholine PC_{20}) may reveal specific miRNAs that may mediate airways hyperresponsiveness epigenetically.

Methacholine PC_{20} is a quantitative marker of airways responsiveness, which is a cardinal feature of asthma and has been tightly linked to exacerbations and other asthma outcomes. This aim investigates the association of circulating miRNA with

methacholine PC_{20} at time of randomization in the Childhood Asthma Management Program (CAMP) [19]. Airway hyperresponsiveness (AHR) in CAMP was an inclusion criterion for the trial; the degree of airway responsiveness has been linked to disease severity [20]. Our hypothesis is that specific miRNAs may be mediating AHR thereby providing unique biologic insights into asthma pathogenesis. Prior work with association of miRNA and pulmonary function testing has been performed revealing biological significance [21]. The association of circulating miRNA and methacholine PC_{20} was assessed. Pathway analysis explored biological significance of these findings. Additionally, a collaborator performed functional validation of miRNA effect in human airway smooth muscle (HASM) cells.

Aim 2: I hypothesize that a panel of circulating miRNA in the CAMP cohort may have good prediction of asthma exacerbation compared to a clinical score yielding a potentially reliable non-invasive biomarker for asthma exacerbation risk

Logistic regression was performed to predict asthma exacerbation from a panel of circulating miRNA and prior asthma clinical score [22]. Exacerbation data from the first year after randomization was used. Based on recent literature [17], it was suspected that a panel of miRNA may have good predictive ability alone or combined with the asthma clinical score for exacerbation.

Aim 3: I hypothesize that machine learning techniques may provide further biological insight and improved prediction of asthma exacerbation. Although logistic regression is more interpretable and may result in good prediction, it is not known if this is the optimal method for prediction with circulating miRNA. A prior study has shown optimization of prediction with a neural network [23]. Both exacerbation (dependent variable) and a panel of miRNA with associated cycle thresholds (independent variable) may have complex non-linear relationships [24], and an artificial neural network may either have improved prediction or biological insight not determined from logistic regression [25] as in **Aim 2.** Additionally, newer algorithms such as XGBoost may also prove suitable for this dataset. XGBoost can be applied to a wide range of

problems, has scalability, can handle data missingness and accept dense or sparse matrices, and includes a regularization term that controls the complexity of models and avoids overfitting [26]. The algorithm also utilizes a unique strategy with tree building compared to random forest. The microRNA dataset is best classified as a dense matrix, and this technique would be suitable to explore the dataset's ability for classification without having to use median imputation for missing values or other pre-processing techniques. However, both neural networks and XGBoost have several hyperparameters that require optimization, which can be computationally expensive. Overall, these strategies may not only yield improved prediction compared to traditional techniques such as logistic regression but also provide biological insights given potential non-linear complexity.

Significance

Prior analyses of circulating miRNA in asthmatic patients have mostly focused on differential expression [27,28], have small sample size and are lacking in quantitative severity measures such as methacholine PC_{20}. For **Aim 1**, an association study of circulating miRNA with methacholine PC_{20}, a proxy of AHR, revealed biomarkers of asthma severity and insight into important epigenetic influences in asthma.

For **Aim 2** and **Aim 3**, both studies are initial explorations in using a panel of circulating miRNA as biomarkers to predict important clinical outcomes in asthma such as exacerbations. These results provided not only insight into the predictive ability of circulating miRNA but also guidance about optimal classification strategies and techniques for future biomarker studies.

These studies are also significant given the large sample size of pediatric asthma patients from a well-characterized cohort, breadth of characterization of circulating miRNA, and use of machine learning techniques.

Chapter 2 - Materials and Methods

A. CAMP Baseline Data, Definition of Asthma Exacerbation, Steroid therapy and tapering description

The CAMP study was a multi-center, randomized, double-blinded clinical trial that explored the safety and efficacy of inhaled budesonide vs. nedocromil vs. placebo in 1041 pediatric patients over a mean follow-up of 4.3 years. The trial design and methodology have been described in detail[18]. Asthma exacerbations were defined as the need for oral corticosteroids (prednisone) for treatment of asthma [19]. The CAMP trial definition of exacerbation, protocol for steroid usage, and tapering description is in the supplemental method section for Chapter 2.

The CAMP Genetics Ancillary Study was approved by each individual study center's Internal Review Board (IRB). Informed consent and assent was obtained from parents and participants, respectively.

B. miRNA Profiling

miRNA profiling in serum from 160 CAMP subjects using TaqMAN miRNA quantitative PCR primers (Life Technologies Megaplex RT Primers, Human Pool Set v3.0, Omaha, NE) containing 754 primers representing 738 unique human miRNAs (miRBase release 21) was performed as previously described [21]. The initial quality control was performed per manufacturer protocol using pre-defined thresholds for amplification scores (> 1.24) and Cq (> 0.80) confidence intervals. All miRNA in this paper are annotated using miRBase release 21 [29] (www.mirbase.org/). The complete miRNA dataset is accessible at the NCBI Gene Expression Omnibus (GEO, http://www.ncbi.nih.gov/geo/) GSE74770.

Analysis of biological replicates was performed in 10% of the samples demonstrating high miRNA-miRNA correlations (rank correlations of > 0.90; data not shown). All subjects were self-identified non-Hispanic Caucasians to limit the effect of race on miRNA expression [30], and subjects were selected from the inhaled corticosteroid arm of the trial. For data analysis, quantile normalization on the detected miRNAs was

performed sample-wise to the mean of the data matrix using MatLab (MathWorks Inc., Natick, MA) function *quantilenorm*. A miRNA was included in the association analysis if expression was present in at least 50 % of samples and the exacerbation prediction analysis if expression was present in at least 70% of subjects. Seven subjects were excluded due to incomplete information regarding steroid burst usage yielding 153 subjects for exacerbation analysis.

C. Data Analyses

1. Association Analysis

Least squares linear regression (both univariate and multivariate) was performed using R [31] to identify miRNA (miR cycle threshold or CT value) associated with the pulmonary function phenotype of interest, $\log_2 PC_{20}$. An association was significant if the 95 % confidence interval for the regression slope (β) does not contain zero. A least squares multivariate linear regression model including miR CT value, age, sex, and height was also calculated for each miRNA. A sensitivity analysis to assess outlier influence, and non-parametric models was also performed. The p-values were corrected using the Benjamini and Hochberg false discovery rate (FDR).

A regulatory network between miRNA and genes was created with usage of Cytoscape (http://www.cytoscape.org/) [32] and CyTargetLinker (http://projects.bigcat.unimaas.nl/cytargetlinker/) [33] with Regulatory Interaction in Network Analysis (RegIN) miRTarBase release 6.1 (http://projects.bigcat.unimaas.nl/cytargetlinker/regins/regins-mirtarbase/) [34] . The Database for Annotation, Visualization and Integrated Discovery (DAVID, Version 6.8 (10/2016), https://david.ncifcrf.gov/home.jsp [35] was used for KEGG [36,37] pathway analysis and gene ontology.

Functional validation of two significant miRNA was performed in human airway smooth muscle (HASM) cells as previously detailed [38]. The cells were transfected with 10nM of either scramble control (AllStars Negative Control siRNA, Qiagen) or miR mimic (Qiagen) using RNAiMax (Life Technology) according to manufacturer's protocol. Seventy-two hours after transfection, cells were trypsinized for 8 minutes and then measured for both cell number and cell size by Moxi Z Cell Analyzer (Orflo). Cell growth was presented as the percentage of cell number relative to scramble control. Average cell diameter (um) was compared in mimic-transfected versus scramble-transfected HASM cells. Data (mean±SE) were obtained from three independent experiments.

2. Clinical score calculation

Using a previously described approach for calculating asthma exacerbation risk [22], a clinical score for predicting pediatric asthma exacerbation was calculated from the baseline history performed during CAMP enrollment in addition to parental survey data [22]. The scoring table is reproduced in **Table E1** (Appendices, Chapter 2 Supplemental Methods and Tables/Figures). Since, leukotriene modifiers were not used in clinical practice at time of study, we capped the total maximum score from the clinical score to be 15 (rather than total of 16).

3. Logistic regression models

Data for asthma exacerbation from the first year after randomization to the inhaled corticosteroid arm was used. Consistent with the National Asthma Education and Prevention Program (NAEPP) guidelines [39], data were divided into dichotomous categories with a total steroid burst count of zero to one and greater than one being assigned as zero and one, respectively. Logistic regression was performed using R [31]. Univariate miRNA and multivariate models adjusting for clinical co-variates (age, sex) were performed. Missing cycle threshold values were replaced sample-wise to the median of the data matrix with the respective column median in order to perform

backwards stepwise selection on significant miRNAs and clinical variables with p-value <0.05. Logistic regression was also performed with the clinical score unadjusted and adjusted for age and sex. Goodness of fit for each model was assessed with the Hosmer-Lemeshow test in the R package ResourceSelection [40]. The likelihood ratio test was performed comparing the different models with R package lmtest [41].

4. Model validation, Receiver Operator Characteristic (ROC) curves and AUROC calculation

R package caret [42] was used to create a 60% training and 40% testing set balancing the class distributions of the outcome, steroid bursts. The R function *predict* was used to predict the outcomes of the models on the testing data set. ROC curves and area under the ROC characteristic (AUROC) were calculated with usage of the pROC [43] and ROCR package [44]. R package cvAUROC [45] was used to perform 10-fold cross validation of AUROCs on the full data set.

5. Identification of miRNA-gene targets and pathway analysis

miRTarBase [46] Release 6.0 (09/15/2015) was used to identify functionally validated gene targets of the three miRNA (miR-206,-146b, and -720) excluding any weak evidence. Pathway analysis with the gene list was performed with use of Database for Annotation, Visualization and Integrated Discovery (DAVID version 6.8, http://david.abcc.ncifcrf.gov/) [47].

6. Additional analyses with neural network and XGBoost algorithms

Further machine learning strategies for classification were applied to the dataset to determine a set of miRNAs that has good prediction of the dichotomized outcome steroid bursts in the first year after randomization. R package, mlr [48] was used for both neural network analysis (based on package neuralnet) and XGBoost (based on package XGBoost).

For neural network analysis, the miRNA dataset cycle threshold cutoff was used as in the logistic regression analyses, and missing values were replaced by the median also

as in the prior analysis. Features were scaled between 0 and 1 to prevent non-convergence issues. Pertinent features were selected based on the Kruskal Test for binary and multiclass classification. For model building, randomized hyperparameter optimization was performed using a) both all features (all the miRNA meeting criteria for inclusion) and b) the top 10 features as determined by the Kruskal test. Both back propagation and resilient backprogation algorithms were performed with the models containing all features and the top 10 features. Parameters including number of hidden layers, threshold, stepmax, and learning rate were determined by randomized hyperparameter selection based on 3-fold cross validation. There were 10 iterations used for tuning. Further analyses with neural networks were not pursued based on these initial results showing suboptimal mean misclassification error.

XGBoost was subsequently explored especially given that it could handle a dense matrix. In contrast to the logistic regression and neural network analyses, missing values were not replaced by the median. Data were not scaled like in the neural network analysis. The “gbtree” (gradient boosted tree) kernel was used. The same dataset split used in the logistic regression model was also implemented for this analysis with a 60/40 % training/testing set. Feature selection was again determined by use of the Kruskal test for binary and multiclass classification. The top 3 features based on this test were used in the model (hsa-miR-206, hsa-miR-409-3p,hsa-miR-30e-3p). Randomized hyperparameter search was performed on the training set with 5-fold cross validation and stratification by the target variable with 100 iterations. The AUROC was subsequently calculated on the testing set data. Given that preliminary analyses did not demonstrate a significant increase of AUROC compared to the logistic regression model, further analyses with this algorithm were not pursued.

Chapter 3 – Results and Discussion for association study with circulating microRNAs and methacholine PC_{20}

A. Results

Study Population

Population characteristics of the 160 CAMP subjects are shown in **Table 1**. The cohort was limited to self-identified non-Hispanic whites due to the significant effects of race on miRNA expression [30]. For the selected individuals, the global characteristics at randomization are representative of the larger CAMP non-Hispanic white cohort (data not shown).

Table 1: Characteristics of CAMP cohort subset

Characteristic	Value (Standard Deviation)
Age – yr	8.8 (2.1)
Sex – no. (%)	Male - 87 (54.4 %)* Female - 73 (45.6%)*
Height – cm	132.7 (13.6)
PC_{20} – mg/mL	1.95 (2.38)
$\log_2(PC_{20})$ – mg/mL	0.06 (1.66)

* The values in parentheses are percentages and not standard deviation

Circulatory miRNA Association with PC_{20}

There were a total of 754 non-housekeeping miRNA mapping to mirBase release 21 on the array, and 155 (20.6%) miRNA were detected in at least 50% of the samples. Eight

microRNAs were significantly associated with PC_{20} (**Table 2**), based on a nominal p-value < 0.05 at a FDR p-value < 0.20. The latter was chosen as a higher cut-off given the nature of this hypothesis generating experiment. Based on prior literature, five of these eight miRNA (63%) had prior evidence of differential expression in human asthma. All associations had a positive slope such that as miR cycle threshold increased so did the PC_{20}; this corresponds to a relationship of increasing miR CT (decreasing miRNA expression) with increasing PC_{20} (decreasing AHR). The strongest association was found with PC_{20} and hsa-miR-296-5p, as shown in **Figure 1**. Sensitivity analysis (**Table E2)** revealed no significant changes in parameters for the models with the exception of non-significance of hsa-miR-30d. Subsequent multivariate analysis including miR CT, age, sex, and height was consistent with the univariate model (**Table E3)**. Nonparametric analysis including both rank-order univariate and multivariate models were also performed and were consistent with the parametric models except for the significance of hsa-miR-451a in the nonparametric model (**Tables E4 and E5)**. Further investigation of miR-30d demonstrated significance in the parametric and non-parametric models with miR-30d cycle threshold characterized by principally having high and low CT values (bimodality) rather than unimodality. This bimodality likely explains non-significance in the sensitivity analysis, while suggesting that miR-30d may still have functional relationship with AHR.

Table 2: Circulatory miRNA Association by Least Squares Linear Regression with methacholine PC_{20} (univariate model, unranked) with detection of miRNA in at least 50 % of samples

miR	Asthma Associated?	miR slope	miR p-value	95 % CI Lower	95 % CI Upper
hsa-miR-296-5p	N	0.460	0.0001*	0.238	0.683
hsa-miR-548b-5p	N	0.328	0.002*	0.126	0.531
hsa-miR-138-5p	Y	0.368	0.003*	0.129	0.608
hsa-miR-16-5p	Y	0.197	0.005*	0.061	0.332
hsa-miR-1227-3p	N	0.327	0.005*	0.100	0.555
hsa-miR-30d-5p	Y	0.201	0.006*	0.060	0.342
hsa-miR-203a-3p	Y	0.203	0.007*	0.057	0.350
hsa-miR-128-3p	Y	0.587	0.012*	0.132	1.042
hsa-miR-942-5p	N	0.242	0.015	0.047	0.436
hsa-miR-451a	N	0.197	0.016	0.037	0.357
hsa-miR-212-3p	N	0.290	0.020	0.046	0.533
hsa-miR-143-3p	N	0.387	0.035	0.028	0.747
hsa-miR-638	Y	0.208	0.048	0.002	0.414

hsa-miR-25-3p	N	0.219	0.049	0.001	0.437

* Significant by FDR adjusted p-value, p < 0.20 cut-off

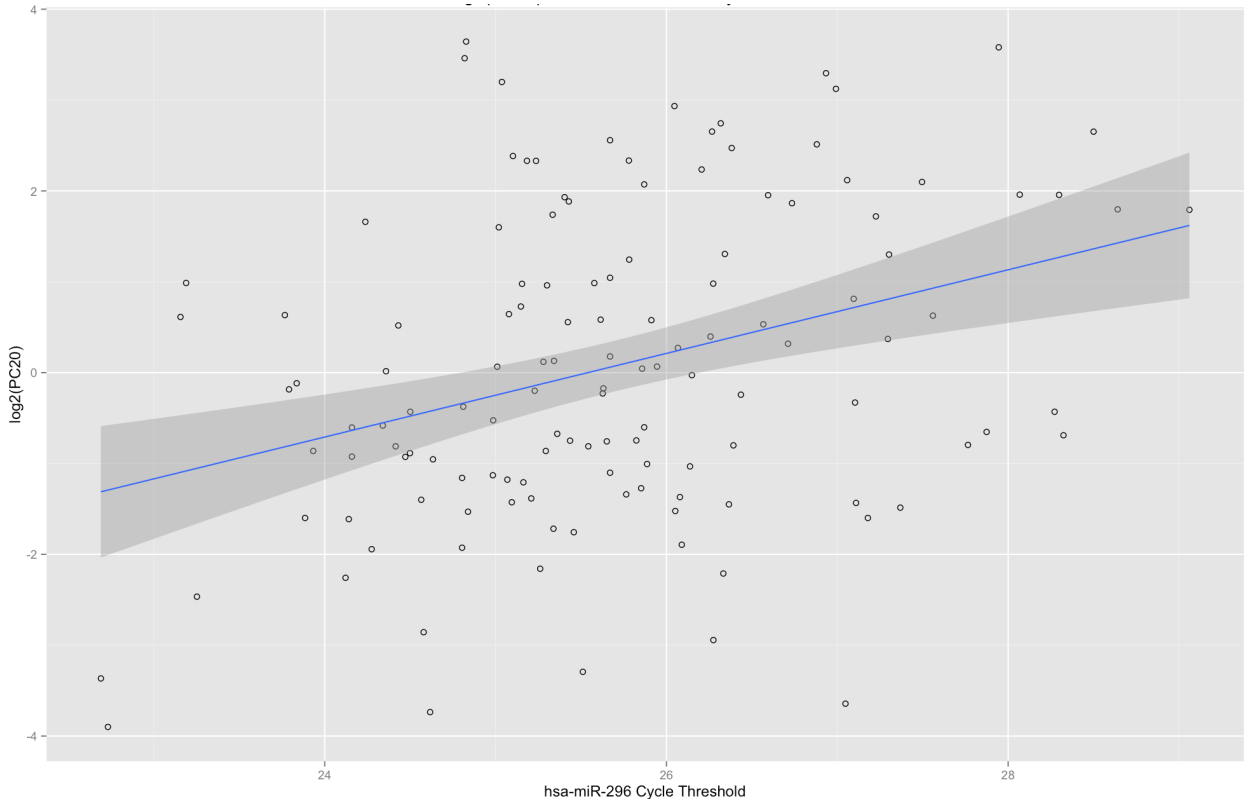


Figure 1. Representative scatter plot of serum miR-296 cycle threshold and log_2PC_{20} in the CAMP Cohort with least squares regression line and 95 % confidence interval.

Pathway and Ontology Analysis

Pathway analysis of the significant miRNAs (**Table E6)** was performed with usage of Cytoscape and CyTargetLinker. The miRNA based on both nominal and FDR p-values were used to generate and create a network with Cytoscape and CyTargetLinker

(**Figure 2**) containing multiple genes. The resultant genes were analyzed with DAVID for KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis with the FoxO and Hippo signaling pathways being the most relevant to asthma (**Table 3, Figure 3a and Figure 3b)**.

Table 3: DAVID Top 10 KEGG Pathway analysis of genes directed from validated miRNA targeting.

Term	Number of Genes in Pathway	Percent of Genes Compared to Total (%)	p-value	Corrected p-value
Signaling pathways regulating pluripotency of stem cells	53	2.0	9.0×10^{-10}	2.6×10^{-7}
Pathways in cancer	103	3.9	1.7×10^{-7}	2.5×10^{-5}
Pancreatic cancer	27	1.0	3.0×10^{-6}	1.1×10^{-4}
FoxO signaling pathway	**44**	**1.7**	**2.8×10^{-6}**	**1.1×10^{-4}**
Hippo signaling pathway	**48**	**1.8**	**2.5×10^{-6}**	**1.2×10^{-4}**

Notes: Threshold for count of 2, EASE 0.1. Table sorted by corrected p-value (Benjamini-Hochberg or false discovery rate). The total number of genes with DAVID ID is 2665.

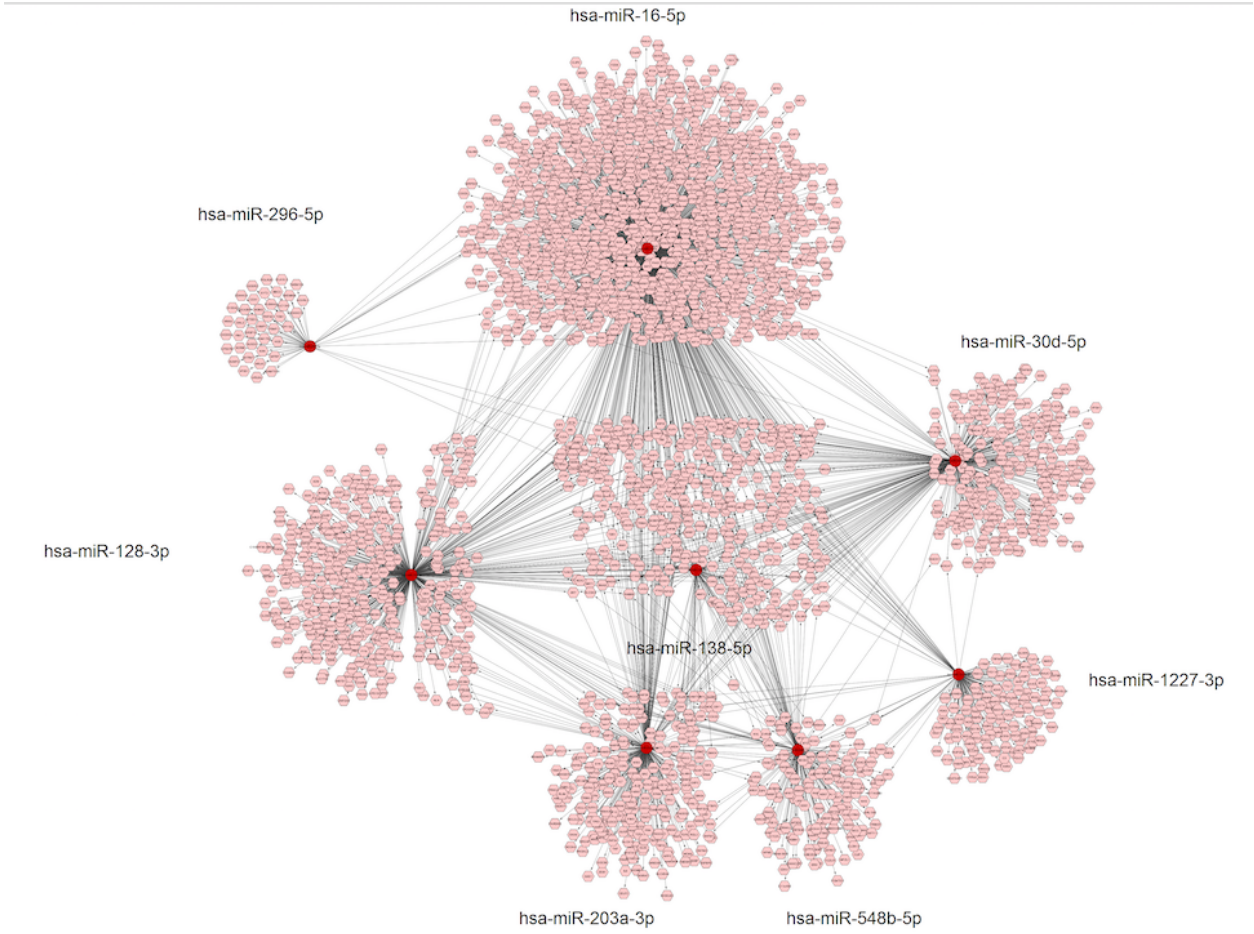


Figure 2. miRNA (red circle) and validated miRNA targeted genes (light magenta circles) predicted by miRTarbase 6.1 in Cytoscape CyTargetLinker with miR-16 having a central connection to other miRNA in the gene network.

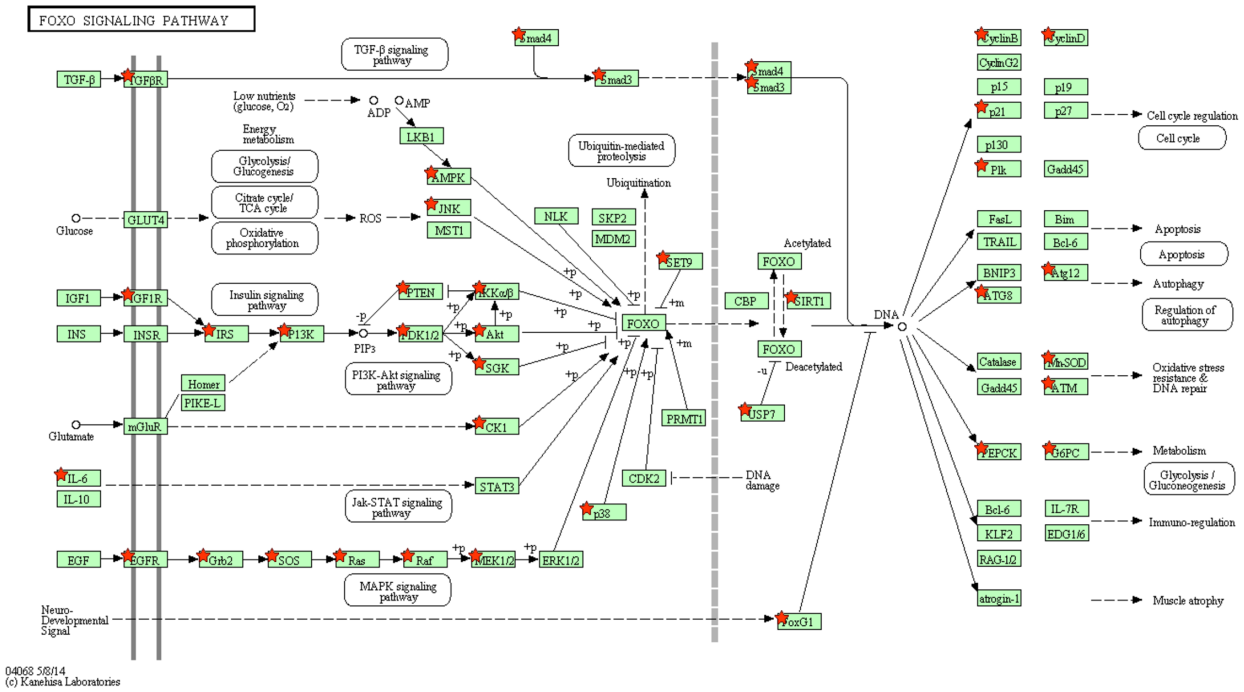


Figure 3a. DAVID KEGG pathway analysis; miR targeted genes (red star) are involved in the FoxO Signaling Pathway**.** Illustration from Kanehisa Laboratories (FoxO signaling pathway, hsa04068), released under Creative Commons Attribution License – Attribution-NonCommercial 4.0 International (CC BY-NC 4.0), 2017, subject to the condition that the original source is acknowledged in the text. For further details, please refer to http://www.kegg.jp/kegg/kegg1.html.

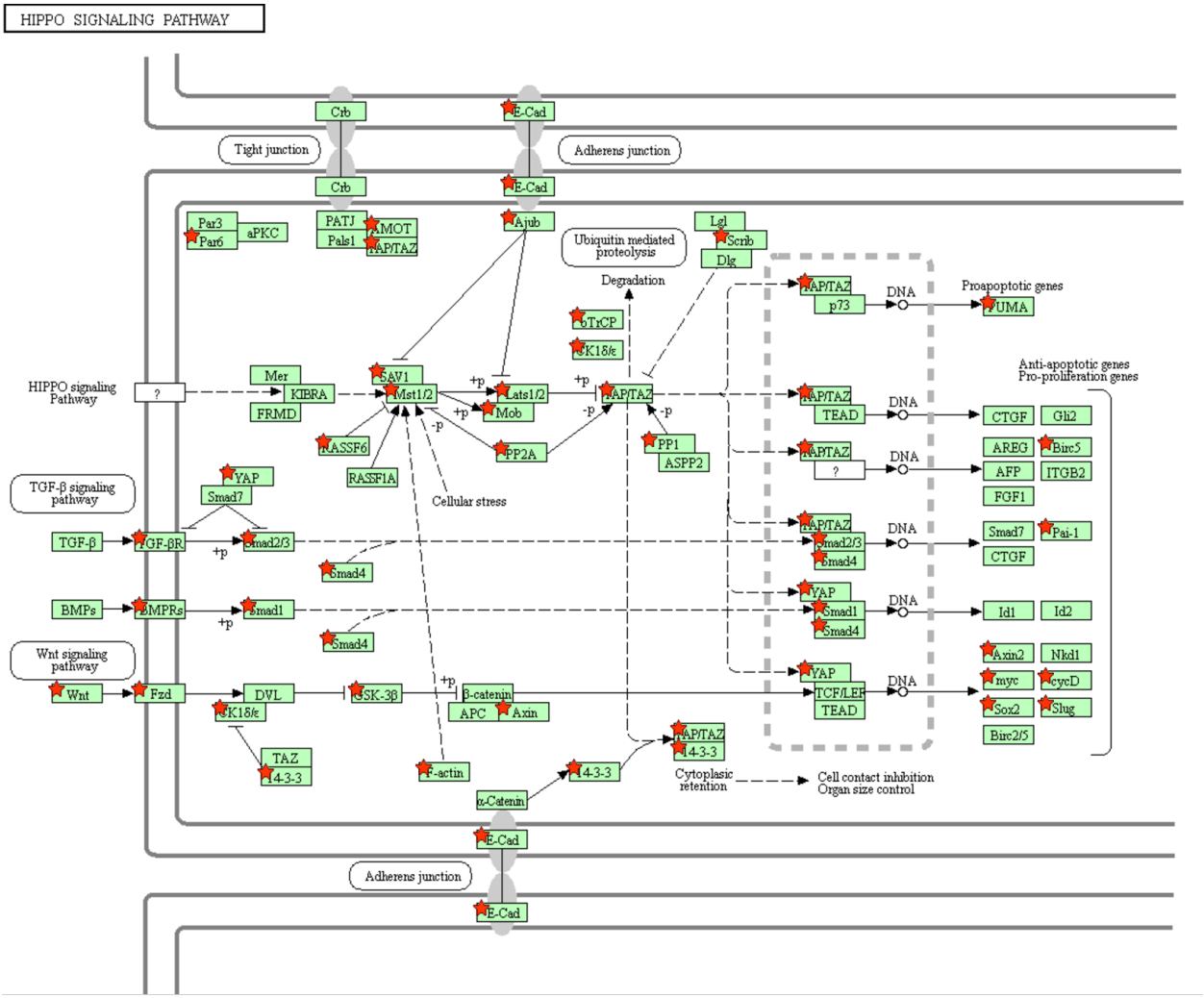


Figure 3b. DAVID KEGG pathway analysis; miR targeted genes (red star) are involved in the Hippo Signaling Pathway**.** Illustration from Kanehisa Laboratories (Hippo signaling pathway, hsa04390), released under Creative Commons Attribution License – Attribution-NonCommercial 4.0 International (CC BY-NC 4.0), 2017, subject to the condition that the original source is acknowledged in the text. For further details, please refer to http://www.kegg.jp/kegg/kegg1.html.

Gene ontology (GOTERM_BP_DIRECT) analysis also revealed functionality of the network related to translation, RNA processing, post-transcriptional regulation of gene expression, ncRNA metabolic process, and other processes (**Table E7).** These functions are consistent with the known actions of miRNA targeting.

Functional validation

Based on our prior miRNA sequencing of human airway smooth muscle cells, [38] of the miRNAs in the PC_{20} network (**Figure 2**), two, miR-16-5p and miR-30d-5p, are significantly expressed. We therefore evaluated the effect of these miRNA on HASM phenotypes using miR-mimics. Mimics of miR-16-5p decreased and miR-30d-5p increased cell growth and average cell diameter, respectively, compared to scramble control (**Figure 4**).

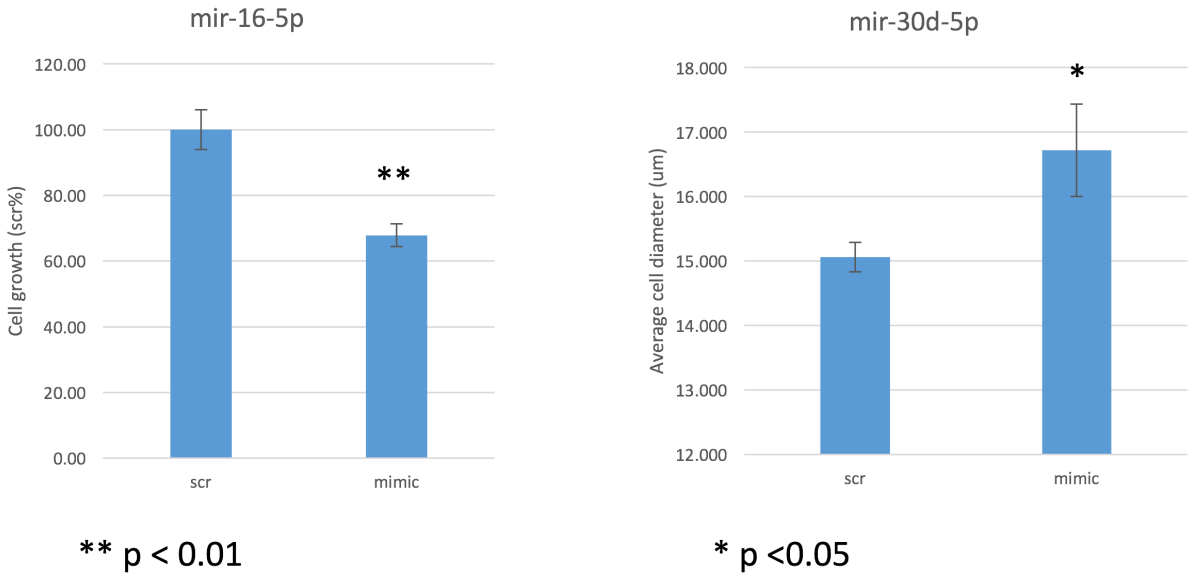


Figure 4. Effect of miR-16-5p and miR-30d-5p on cell growth and average cell diameter, respectively compared to scramble control. HASM cells were transfected with 10 nM of either scramble control or miR-16-5p mimic (left panel; or miR-30d-5p in right panel). Seventy-two hours after transfection, cells were trypsinized and measured for both cell number and cell size by Moxi Z Cell Analyzer. Cell growth was presented as the percentage of cell number relative to the scramble control. Average cell size (um in diameter) was compared in mimic-transfected versus scramble-transfected HASM cells. Data were obtained from three independent experiments and were contributed by Drs. Maoyun Sun and Quan Lu.

Discussion

In this study, we examined serum samples from 160 CAMP asthmatics and found 8 miRNA significantly associated with PC_{20}, a defining measure of airways hyperresponsiveness. Based on prior literature, five of the eight miRNA (63%) had evidence of differential expression related to human asthma, but not $PC_{20,}$ with a good portion of these in case-control studies of human bronchial epithelial cells. Three novel miRNAs were identified, including our strongest association, miR-296-5p. Pathway analysis of the miRNA targets implicates effects of both the Hippo and FoxO signaling pathways with both pathways implicated in airways hyperresponsiveness [49,50]. Lastly, functional validation demonstrated that miR-16-5p resulted in decreased airway smooth muscle cell growth and miR-30d-5p increased airway smooth muscle cell size compared to scramble controls.

Our most significant association was found with hsa-miR-296-5p (**Table 2**). There are no previous reports in the literature regarding this miRNA in association with asthma. miR-296 targets *IKBKE*, which is involved in signaling pathways including Toll-like receptor signaling and signal transduction prompting apoptosis [51]. *IKBKE* is highly expressed in immune cells and is a known target of the NFκB gene [52]. The NFκB pathway's involvement in asthma and inflammation has been well described in the literature [53], and includes modulation of AHR in allergen challenged mice [54]. Moreover, *IKBKE* itself is a known therapeutic target for asthma, with IKBKE targeting demonstrating significant attenuation of airways responsiveness and inflammation in a

murine model of asthma [55]. Therefore, miR-296 may attenuate immune response and could modulate AHR via the NFκB pathway.

miR-16-5p was also significant in our study and differential expression of this miRNA in asthmatic airway cells has been reported [56]. Expression profiling of human airway biopsies has showed miR-16 to be highly expressed, leading to the hypothesis that miR-16 along with other miRNAs may have a significant influence on gene expression in the airways [57]. Our network analysis demonstrated that miR-16 plays a key role as the central hub, both interacting with other miRNAs and mediating expression of dozens of genes (**Figure 2**). Thus, miR-16 appears to play a notable role in the modulation of genes influencing airways hyperresponsiveness in asthma. In addition to its central effect on downstream gene expression, miR-16 mimics result in decreased airway smooth muscle growth. While the exact significance of this finding is unknown, prior work focused on small airway cell layers suggests that differential growth between layers may mediate different effects on airway buckling [58].

As mentioned, several of our other AHR associated miRNA, including hsa-miR-30d, -128, -138, and -203a, have been detected in studies involving human airway cells of asthmatics [28]. The association of hsa-miR-203 has been validated in epithelial cells from a small number of asthmatics and healthy subjects with identification of the top-ranked predicted target, aquaporin gene (AQp4). In turn, the expression of AQp4 was subsequently noted to be significantly higher in asthmatic cells [28]. Other studies have shown up-regulation of miR-203 in serum of children with atopic dermatitis and increased IgE level [59] in addition to airway epithelial cell apoptosis [60]. Thus miR-203

may indirectly affect airways responsiveness via an inflammatory mechanism. In contrast, our work demonstrates that miR-30d-5p increases average HASM cell diameter compared to scramble controls. Increased airway smooth muscle cell size can result in both further mechanical airway narrowing in addition to increased contribution of inflammatory mediators [61]. Increase in airway smooth muscle tissue mass related to both hypertrophy and hyperplasia has been noted a major driver of airway narrowing and thus AHR in asthmatics [62]. It is very likely that miRNA act via increases in ASM cell size/diameter and thus, mechanistically may directly cause increased AHR (decreased PC20).

Focusing on validated miRNA targets, pathway analysis from our associated miRNAs was notable for multiple genes in both the FoxO and Hippo signaling pathways (**Figure 3a and 3b**). For the former pathway, a mouse experiment showed alternative activation of alveolar macrophages with resultant type 2 allergic airway inflammation with subsequent airway remodeling [49]. For the latter pathway, it has been shown that it is a notable regulatory pathway with versatile function including a key gene (Yes-associated protein or YAP) implicated in airway smooth muscle hyperplasia [50]. Both of these pathways have a plausible link to the phenotype of airways hyperresponsiveness. As noted above, miR-16 also appears to be a central hub in our serum microRNA network and may work in concert with other miRNA to modulate immune pathways and subsequently AHR. Functional validation would be needed for further elucidation of possible molecular mechanisms between miRNAs and asthma related to this pathway. Lastly, gene ontology analysis (**Table E7**) demonstrated processes such as RNA

processing, post-transcriptional regulation of gene expression, and other likely putative effects of miRNAs.

This study has several strengths including a large sample size of pediatric asthma patients from the CAMP cohort, a large number of interrogated miRNAs, validation of prior associations in the literature with our reported miRNA findings, and subsequent functional validation of miRNA in HASM. The large sample size and number of interrogated miRNAs provides a good breadth of characterization and power to detect associations in light of lower starting concentrations of miRNA in the circulation. Additionally, the CAMP cohort was clinically well characterized with standard methodologies including methacholine challenge testing, which should minimize potential for measurement error. Analysis of biological replicates as discussed in the methods section also showed high miRNA-miRNA correlations. Although the CAMP serums were stored for years prior to this interrogation, prior studies have shown the stored samples can result in reliable miRNA concentrations months to years later [63]. Lastly, miRNA targeting is an imprecise science with new associations being discovered on a regular basis. However, our study used miRTarBase (validated miRNA-target interaction), which assesses only functionally annotated miRNAs, lending functional credence to our network and pathway analyses; this was enhanced by our functional studies in HASM cells

In summary, this study detected eight circulating miRNAs associated with PC_{20} in a pediatric asthma population with mild-moderate persistent asthma. These miRNA appear to be associated with individual and pathway evidence of immune modulation

that could affect AHR; complementary functional validation of miR-16-5p and miR-30d-5p in HASM demonstrate effects on cell growth and diameter, respectively. The majority of these miRNAs had been associated with asthma in prior studies. Nonetheless, the most significant association was a novel association with miR-296, and this miRNA may be a viable serum biomarker for altered immunity and AHR in pediatric asthmatic patients.

Further study of our PC_{20} associated miRNAs, both in terms of external validation and additional functional mechanisms, may provide insight into epigenetic influences in asthma pathobiology and have clinical implications such as risk assessment and treatment responses. Given that miRNA can therapeutically decrease airways responsiveness in murine models of asthma [64-66], future work may also yield novel therapeutic approaches to targeting asthma via miRNA modulation of AHR.

Chapter 4 – Results and Discussion for prediction of asthma exacerbation with panel of circulating microRNAs compared to an asthma exacerbation clinical score

Results

Study Population: Population characteristics of the 153 CAMP subjects stratified by asthma exacerbation status are shown in **Table 4.** About 25 % (38 of 153) of the subjects experienced an exacerbation during the first year of the trial, despite being randomized to inhaled corticosteroids. The cohort was restricted to self-identified non-Hispanic whites to eliminate confounding from race on miRNA expression [30].

Table 4: Cohort Characteristics stratified by asthma exacerbation status

No. Subjects	No Exacerbation N= 115	Exacerbation # N = 38	
Characteristic	Mean (SD^)	Mean (SD^)	p-value
Age - years	8.9 (2.0)	8.9 (2.2)	0.93
Male sex - %	54%	61%	0.48*
Height – cm	133.2 (13.4)	132.8 (13.8)	0.86
BMI - percentile	61.5 (28.3)	63.1 (26.4)	0.75
FEV1 pre-bronchodilator - L	1.7 (0.5)	1.6 (0.5)	0.71
# of steroid bursts in one year	0.3 (0.4)	3.2 (1.7)	$1.4x10^{-12}$
Clinical Asthma Exacerbation Score (# of subjects)	Low 13 Average 43 High 59	Low 1 Average 10 High 27	0.08^

^ SD = Standard Deviation. If not present, value is absolute value and NOT mean

Exacerbation status: Participant who received 0-1 steroid burst in one year were classified as “no exacerbations.” Two or more steroid bursts were classified as “exacerbation.”

* Chi-square test. ^ Fisher's Exact test. Others p-values calculated by two-sided, independent t-test not assuming equal variance

miR Association and Prediction of Asthma Exacerbation: Twelve of the 125 interrogated miRNA were associated with exacerbation for pediatric asthma exacerbation within the first year following randomization (**Table 5**). The strongest miRNA association for asthma exacerbation was miR-206 with an OR 0.60 (95% CI: 0.42-0.83). Both the miRNA (**Figure 5**, Panel A) and clinical asthma exacerbation score (**Figure 5**, Panel B) logistic functions are shown in **Figure 5**. All the miRNA associations demonstrated odds ratio less than one indicating decrease risk of asthma exacerbation with increase in miRNA CT (i.e. decrease in circulating miRNA abundance). Overall, each doubling of expression of the 12 associated miRNAs was associated with between a 25 to 67% increase in risk of exacerbations. Following our initial univariate models, we formulated a prediction model from the associated miRNAs using stepwise (backwards) logistic regression. Following selection, a three miRNA model demonstrated best fit – miR-206, miR-146b, and miR-720 (**Table 6**).

Table 5: Univariate (unadjusted) logistic model for miRNA with unknown values replaced by median

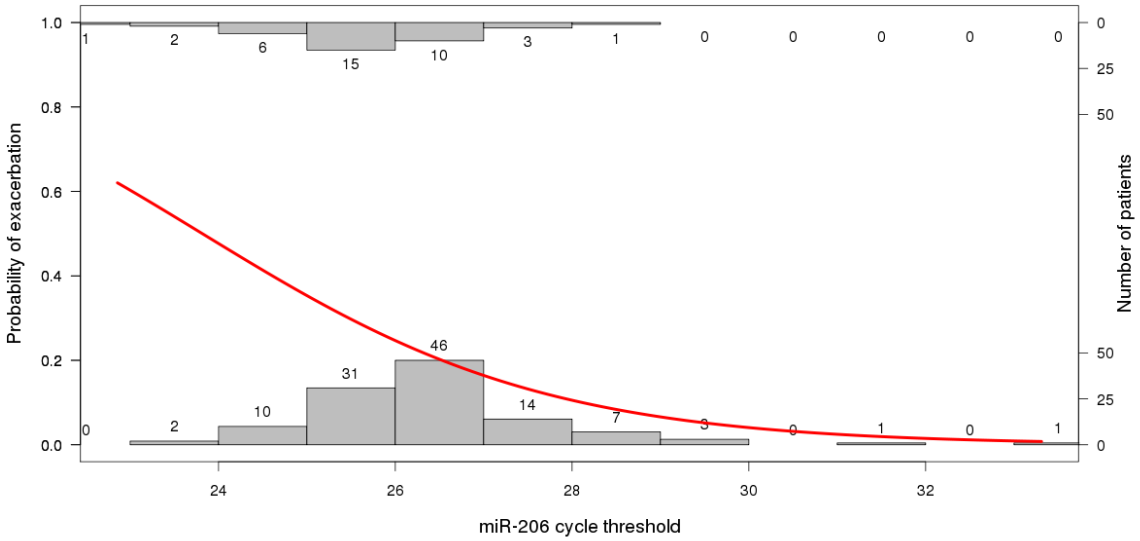
miRNA	Odds ratio (OR)	95 % Confidence Interval	p-value
miR-206	0.60	0.42-0.83	0.004
miR-146b-5p	0.66	0.48-0.89	0.007
miR-222-3p	0.70	0.52-0.93	0.02
miR-409-3p	0.73	0.56-0.95	0.02
miR-223-5p	0.62	0.40-0.92	0.02
miR-126-5p	0.68	0.48-0.93	0.03
miR-339-3p	0.72	0.53-0.96	0.03
miR-30e-3p	0.70	0.49-0.95	0.03
miR-126-3p	0.74	0.56-0.96	0.03
miR-342-3p	0.80	0.64-0.98	0.04
miR-454-3p	0.77	0.60-0.98	0.04
miR-720	0.71	0.50-0.98	0.046

Table 6: Summary of Logistic Regression Models with coefficient odds ratio with 95 % confidence intervals and model measures

Variable	miR model[#]	Clinical Model	miR + Clinical model
hsa-miR-206[&]	0.64**[$] (0.45,0.89)	—	0.65** (0.44, 0.92)
hsa-miR-146b	0.72** (0.52,0.98)	—	0.67** (0.47,0.93)
hsa-miR-720	0.75 (0.51, 1.1)	—	0.70* (0.46,1.03)
Clinical score	—	1.36*** (1.14,1.64)	1.38*** (1.15,1.69)
Age	—	1.01 (0.84,1.23)	0.97 (0.80,1.20)
Sex	—	0.79 (0.36,1.73)	0.70 (0.29,1.65)
Observations	153	153	153
Log Likelihood	-76.8	-79.5	-69.9
AIC[+]	161.6	167	153.7
Hosmer–Lemeshow Goodness of Fit Test^	X^2= 9.15 p-value = 0.33	X^2= 8.45 p-value = 0.39	X^2= 7.07 p-value = 0.53

The miRNA model was determined by backwards stepwise selection. The clinical score model was adjusted for age and sex as per original publication.
& Each of the model coefficients presents the odds ratio and the 95 % confidence interval in parenthesis
$ P-values are denoted by an asterisk (*) with: * p <0.1, ** p < 0.05, *** p < 0.01
+ Akaike information criterion
^ Ten bins used to calculate quantiles with df = 8

PANEL A



PANEL B

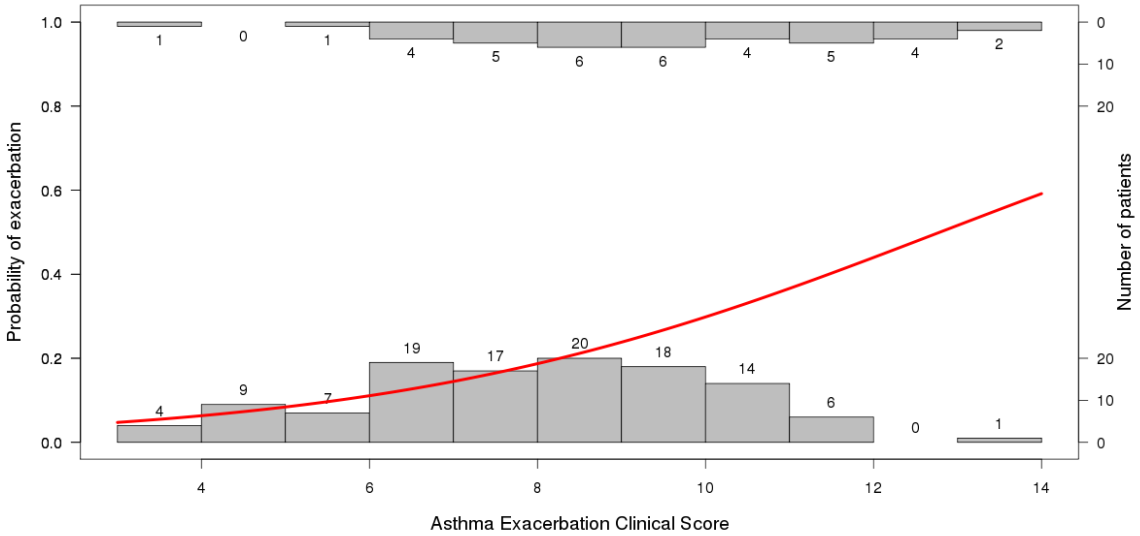


Figure 5: Unadjusted logistic regression models for miR-206 (Panel A) and asthma exacerbation clinical score (Panel B). The histograms represent the observed data with the bottom axis representing the miRNA cycle threshold and asthma exacerbation score for panel A and B, respectively. The number of patients is represented on the right axis. The top histograms represent the number of subjects that had an exacerbation and the bottom histograms represent the number of subjects that did not have an exacerbation at the respective miR-206 and asthma exacerbation clinical score, respectively. The red line represents the unadjusted logistic regression function with probability of exacerbation on the left axis. For instance in panel A, as the miR-206 cycle threshold increases (less detection in the blood), the risk for asthma exacerbation decreases. Conversely in panel B, as the asthma exacerbation score increases, the probability of asthma exacerbation (left axis) increased.

miR vs. Clinical Model. The model for the clinical score demonstrated a positive coefficient with resultant odds ratio 1.36 [95 % CI: 1.14 – 1.64] indicating an increased risk of asthma exacerbation with increase in clinical score. The miRNA, clinical, and combined miRNA-clinical models are summarized in **Table 6.** Likelihood ratio testing was performed indicating that the training set combined miR-clinical model were superior to the clinical model (X^2 11.7, P-value 0.009).

Model prediction was assessed via ROC curves generated for the miRNA, clinical score, and combined miRNA-clinical model training set data (**Figure 6**).

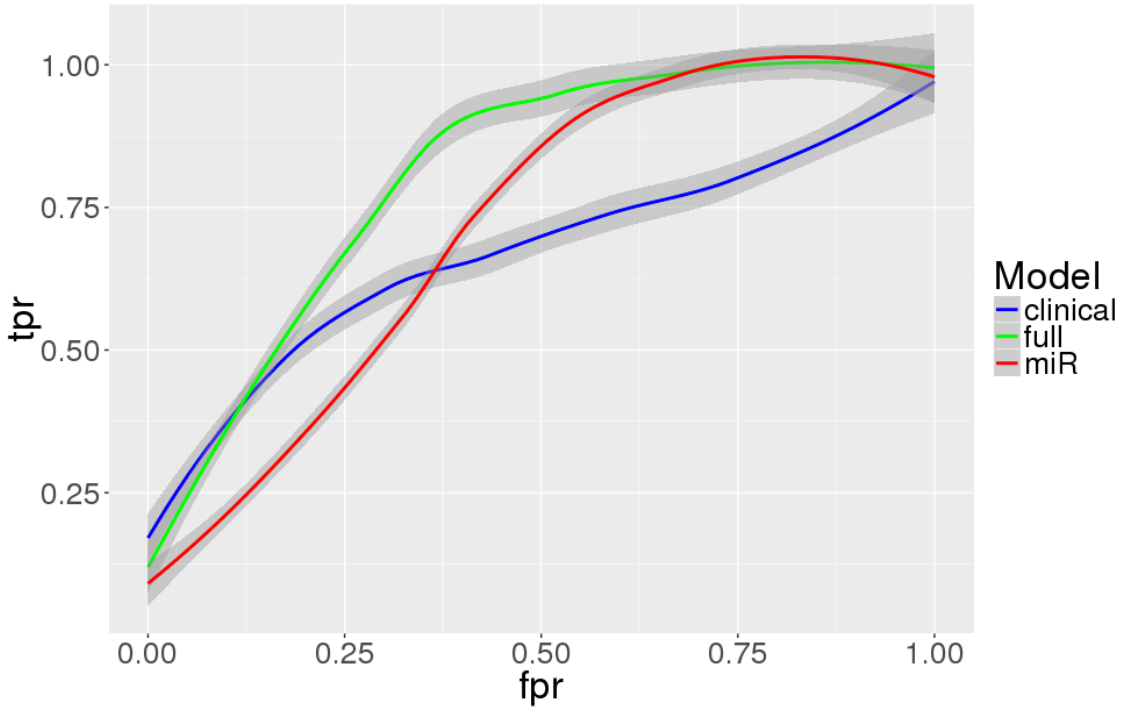


Figure 6: Comparison of AUROC between the asthma exacerbation clinical score model (blue line), miRNA (red line) model, and full (miRNA and clinical score, green line) models in the testing set data. AUROCs were 0.671, 0.714, and 0.807, respectively. A linear LOESS smoothing function was applied with 95 % confidence interval shown. Legend: Y-axis: tpr = true positive rate, X-axis: fpr = false positive rate

For the testing set, the combined miRNA-clinical score model demonstrated good predictive power with an AUROC of 0.81; this was significantly better than either the clinical model (AUROC 0.67) or miRNA model (AUROC 0.71) alone. Ten-fold cross-validation AUROC was performed on the full data set showing that the estimated AUROC for the combined model retained good predictive value and remained higher than both miRNA and clinical model (**Appendices, Chapter 4, Table E8).**

Gene Ontology and pathway analysis: The top asthma-related pathways in the DAVID Biocarta Pathway Analysis were Inactivation of Gsk3 by AKT causes accumulation of b-catenin in alveolar macrophages (FDR p-value 1.7 x 10^{-2}) and NF-κβ signaling pathway (FDR p-value 8.0 x 10^{-2}) (**Table E9**). Multiple miRNA-targeted genes

were in the Inactivation of Gsk3 by AKT causes accumulation of b-catenin in alveolar macrophages pathway (**Supplemental Figure 1**). Both of these pathways have significance in asthma. Gene ontology was performed and can help define concepts related to gene function and aspects of biological processes. Gene ontology was most significant for positive regulation of fibroblast proliferation (FDR p-value 1.5×10^{-2}) (**Table E10**).

Discussion

In this study, we examined baseline circulating miRNA in childhood asthmatics prior to treatment with inhaled corticosteroids (ICS) to predict exacerbations in the subsequent year. We noted that 12 miRNAs were significantly associated with future exacerbations, with each doubling of expression of these miRNAs associated with a 25-67% increase in risk of exacerbations. When combined, three miRNAs by themselves provided comparable predictive power to an established clinical model of exacerbations. Moreover, when the miRNAs were combined with the clinical model, significant increases in the ability to predict future exacerbations were noted, with an overall AUROC of 0.81. To our knowledge, this study is the first to examine prediction of asthma exacerbations with miRNAs.

Notably, all of the subjects within the current study were randomized to inhaled corticosteroids as part of a clinical trial cohort. Therefore, while our findings may apply broadly to asthma exacerbations as a whole, this information may also have pharmacogenomics implications via the identification of subjects that experience exacerbations despite therapy with inhaled corticosteroids. Therefore, validation of these findings may signify subjects that may benefit from alternate, or additional, therapies.

Our top miRNA, hsa-miR-206, was significantly associated with subsequent asthma exacerbations with a notable odds ratio (OR) of 0·60 (95% CI: 0.42-0.83) (**Table 5**). Given that the distribution of expression for this miR is very broad (spanning multiple cycle thresholds, as in **Figure 5**, Panel A), this suggests that the reported OR for the

individual miRNAs are likely conservative. **Figure 5** demonstrates the difference between the miRNA (Panel A) and clinical asthma exacerbation score (Panel B) logistic regression function with fairly dispersed values for exacerbation status compared to a division at miR-206 cycle threshold around 26. This visualization suggests miRNA has a better discrimination between exacerbation and non-exacerbation status compared to the clinical asthma exacerbation score.

Stepwise selection of significant miRNAs resulted in a three miRNA model (hsa-miR-206, -146b, and -720) that was subsequently compared to and combined with an asthma exacerbation clinical score model (**Table 6)**. The combined miR-clinical model demonstrated very good to excellent predictive power (**Table 6**), with an AUROC to classify exacerbation vs. non-exacerbation of 0.81. Notably, this AUROC is significantly higher than prior clinical studies focused solely on the asthma exacerbation clinical score [22] with an AUROC of 0.75 for a Costa Rican cohort and AUROC 0.69 for the CAMP cohort and Childhood Asthma Control Test (C-ACT) with AUROC 0.72 [67]. The AUROC is also improved compared to other biomarkers including addition of our prior study of GWAS SNPs from the same cohort [7]. The combined miR-clinical model far exceeded the predictive capability for exacerbation status compared to the miRNA or clinical score model alone. These data support that optimal prediction for personalized medicine are likely to come from a combination of ‘omics and clinical variables.

The miRNAs used to build our predictive model overlaps with biologically relevant miRNAs related to asthma pathobiology. miR-146b, along with miR-146a, are negative regulators of inflammatory gene expression in lung tissue [68]. A murine model study of acute and chronic asthma demonstrated consistent upregulation of miR-146b, which is also expressed by leukocytes and has been shown as a negative regulator NF-κB in human breast cancer cells [69]. This could potentially explain why circulating miR-146b may be a viable biomarker. In eosinophilic esophagitis, miR-146b has been found both in esophageal biopsies and differentially regulated in the plasma [70]. Altered expression in the airway wall of the other two miRNAs in our predictive model, miR-206 and miR-720, were noted in a mouse model of childhood allergic asthma [71]. Moreover miR-206

has been shown to be involved in airway smooth muscle (ASM) innervation [72], thereby enhancing the mechanistic significance of our model. Integrative miRNA studies would be needed for direct elucidation of interactions of our miRNAs as they relate to asthma pathobiology. Pathway analysis (**Table E9**) also had borderline significance for the NF-κB pathway, which correlates to the aforementioned miR-146b regulatory role. Genes affected by the miRNA were also involved in the inactivation of Gsk3 by AKT causes accumulation of b-catenin alveolar macrophages pathway (**Supplemental Figure 1, Table E9)**. Inactivation of Gsk3 has been studied in a murine model and is associated with ASM hypertrophy and possibly linked to asthmatic airway remodeling [73]. Overall, further study of these miRNA may suggest a functional approach to small RNA directed therapies toward the treatment or prevention of asthma exacerbations.

This study has several strengths including a large number of interrogated miRNAs, large sample size of pediatric asthma patients from the CAMP cohort, biologically significant miRNAs found in modeling, and comparison to a prior validated clinical score. The CAMP cohort was clinically well characterized notably for both identification and protocol-based treatment of asthma exacerbations, which should reduce measurement error. Replicate analysis discussed in the methods section showed high miRNA-miRNA correlations. While the primary CAMP trial recruitment occurred several years ago, circulating miRNAs have been demonstrated to be stable over the course of many years [74,75]. While we do not yet have independent replication, the maintenance of high AUROCs in our cross-validation analyses, support good potential for generalizability.

In summary, our results are promising for the translation of circulating miRNA to predict clinical outcomes related to asthma and are consistent with prior cited literature about usage of circulating miRNA as potential biomarkers, predictors, or markers of treatment response. miRNA as biomarkers for the *a priori* prediction of asthma exacerbations may be particularly salient, given the substantial morbidity and health care costs associated with exacerbations. Additionally, since this particular study restricted subjects to inhaled corticosteroid therapy, its findings may have direct pharmacogenomic relevance in the

prediction of which participants respond to ICS therapy. Further study of miRNA alone or in concert with other genomic and epigenomic markers may reveal additional predictive power for asthma risk assessment and even treatment responses.

Chapter 5 – Results and Discussion for prediction of asthma exacerbation with panel of circulating microRNAs with additional machine learning techniques

Results

Study Population:
The same subjects from chapter 4 were used for this analysis. Population characteristics of the 153 CAMP subjects stratified by asthma exacerbation status are shown in Chapter 4, **Table 4**. Again, approximately 25 % of subjects (38 out of 153) had an asthma exacerbation during the first year of the trial after being randomized to inhaled corticosteroids.

Feature Selection:
The Kruskal test for binary and multiclass classification tasks from the mlr package[48] was used for this exploratory analysis given its non-parametric nature and non-specificity to a particular classification method. This method "applies a Kruskal-Wallis rank sum test of the null hypothesis that the location parameters of the distribution of a feature are the same in each class and considers the test statistic as a variable importance measure" per mlr documentation [48]. As in chapter 4, miRNA were only included if expression was present in at least 70% of subjects. For the neural network, features had undergone median imputation for missing values, were scaled between 0 and 1, and subsequently underwent feature selection. **Figure 7** demonstrates the relative importance of the top twenty features as determined by the Kruskal test for the neural network feature selection. The top miRNA based on the Kruskal test was miR-206, which was also the top miRNA in the logistic regression model (Chapter 4, **Table 5**) where feature selection was based on univariate significance. For XGBoost, the features did not undergo scaling and imputation as these procedures are not needed for

this algorithm. The top 3 features for the XGBoost model as determined by the Kruskal test for binary and multiclass classification were hsa-miR-206, hsa-miR-409-3p, and hsa-miR-30e-3p.

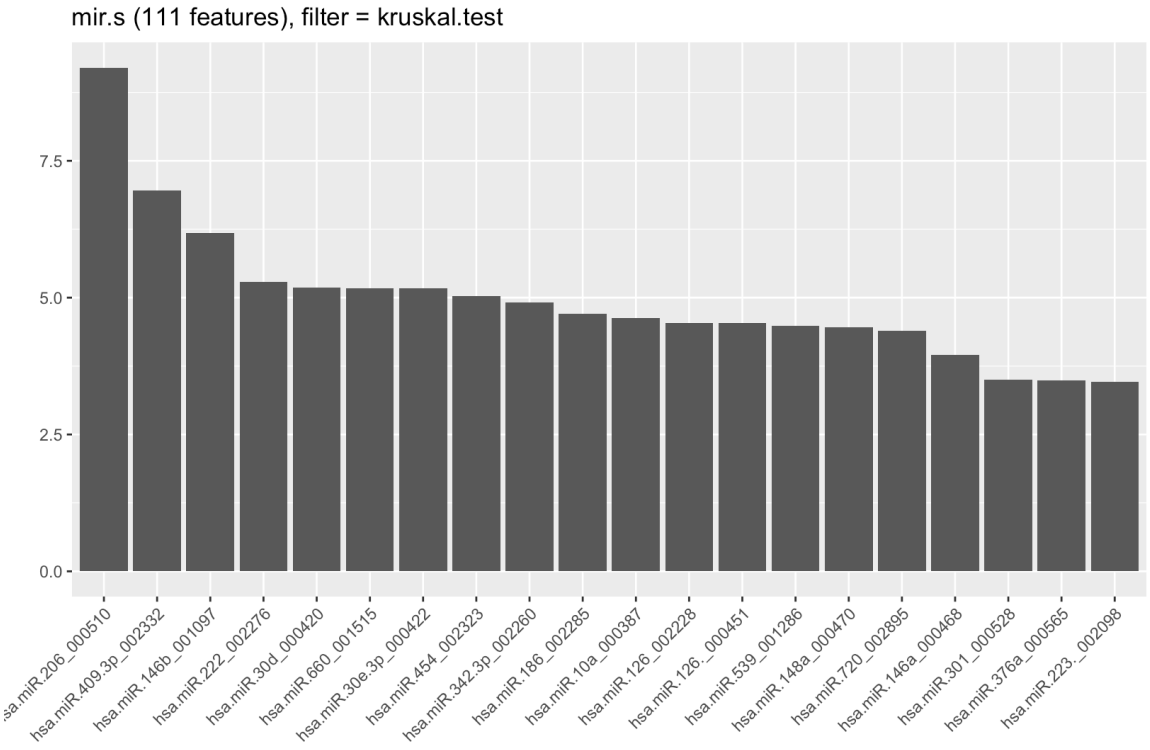


Figure 7: Neural Network Feature selection. Top features in the miRNA dataset as determined by the Kruskal Test for binary and multiclass classification. The y-axis contains the value of the test statistic as a variable of importance measure as described in the test, and the x-axis contains the feature names arranged in ascending order based on the test statistic.

Neural Network Hyperparameter Optimization and Modeling

Four total neural networks were built with differing features and algorithms and underwent randomized hyperparameter optimization as in **Table 7**. Model 1 had the smallest mean misclassification error with the back propagation algorithm and all features. The worst performance was the model with the resilient back propagation algorithm and all features. Model 2 and 4 contained the top 10 features as determined by the Kruskal Test and were ran with the back propagation and resilient back propagation algorithms, respectively. These models have a similar mean

misclassification error. Further analyses were not performed given these initial results yielded performance near the no information rate of the model.

Table 7: Randomized Hyperparameter Optimization with neuralnet package for the miRNA dataset

Characteristic	Model 1	Model 2	Model 3	Model 4
Features	All miRNA	Top 10 Features	All miRNA	Top 10 Features
Algorithm	Backprop	Backprop	Rprop+	Rprop+
Hidden	1	12	2	18
Threshold	0.673	0.0822	0.0755	0.0643
Stepmax	818106266	473288965	320425282	960710231
Learning Rate	0.0366	0.112	0.0731	0.0883
Mmce.test.mean	0.248	0.306	0.32	0.294

Notes:
Cross-validation x 3, 10 iterations for tuning
Features indicate either all the features in the microRNA dataset ("all miRNA") or the top 10 features as determined by the Kruskal test for binary and multiclass classification
Algorithm refers to "backprop" and "Rprop+", which is backpropagation and resilient backpropagation, respectively.
Hidden refers to the number of hidden layers in the neural network
Stepmax and learning rate are defined parameters in the neuralnet package.
Mmce.test.mean is the mean misclassification error defined as the mean of response not equal to truth

XGBoost Hyperparameter Optimization and Modeling

The XGBoost modeling was performed with the top 3 features determined by the Kruskal test using the gradient boosted tree (gbtree) kernel. Modeling was performed on the training set data with randomized tuning of relevant hyperparameters as demonstrated in **Table 8**. Model prediction was assessed via AUROC. The model was subsequently applied to the testing set data with a resultant AUROC of 0.69. This AUROC was close to the AUROC of 0.714 determined by a panel of three miRNA in the logistic regression model (Chapter 4, **Figure 6**) although the only one of the features (miR-206) between the logistic regression and XGBoost was the same.

Table 8: Randomized Parameter Optimization for XGBoost

Parameter	Tuning Value	Constraint
Kernel: gbtree	Not applicable	---
Max depth	3	3 - 10
Lambda	0.477	0.05 - 0.5
eta	0.27	0.01 - 0.5
subsample	0.775	0.5 - 1
Min_child_weight	4.94	2 - 10
Colsample_bytree	0.57	0.50 - 0.80
AUROC	0.69	0.5 - 1

Notes:
Kernel was gbtree (gradient boosted tree) in the R package, XGBoost
Data were divided into 60 % training and 40 % testing set.
Randomized hyperparameter search was performed on the testing set data with 5-fold cross validation and stratification by the target variable with 100 iterations.
AUROC is the area under the receiving operator curve for the testing set

Discussion

In chapter 4, a panel of baseline circulating miRNA prior to randomization to the inhaled corticosteroid arm in the CAMP cohort was used to predict asthma exacerbations in the subsequent year alone and in conjunction with an asthma clinical score. While logistic regression results are interpretable and provided good prediction based on AUROC, exploration with other machine learning techniques related to prediction using circulating miRNA has been limited with recent pulmonary publications consisting of neural network [23] and random forest[17]. In the exploratory analyses in this chapter, two techniques including neural networks (both back propagation and resilient back propagation algorithms) and XGBoost were explored. Overall, the neural network technique did not yield good accuracy; whereas, the XGBoost algorithm produced an AUROC slightly lower than the logistic regression miRNA model.

For the neural network, the lowest mean misclassification error (mmce) in **Table 7** was 0.248 in model 1. However in light of the dataset having class imbalance between exacerbation (~ 25 %) and no exacerbation (~ 75 %), the mmce is overall not excellent regardless of features and algorithm. Further analyses were not performed given these initial results but may be explored in the future on a larger dataset. The analyses do not support the hypotheses about neural networks providing further biological insights and exploring complex non-linear relationships [24] at least in this miRNA dataset.

The study has several strengths as mentioned in Chapter 4 such as a large number of interrogated miRNA and large sample size of pediatric asthma patients from the CAMP Cohort. Limitations to the neural network analyses include lack of alternative feature selection methods described by Olden [76] as this method is difficult to implement into the mlr package pipeline. The network parameters may have not been optimized given the low number of iterations for randomized hyperparameter optimization.

For XGBoost, the testing set AUROC with a panel of 3 miRNA (including one miRNA that overlaps with the panel for the logistic regression model) was 0.69 (**Table 8**), which was slightly less than the AUROC of the logistic regression model (Chapter 4, **Figure 6**). These two results using both different modeling approaches and data pre-processing methods (e.g. median imputation vs. dense matrix input for logistic regression and XGBoost, respectively) support robust prediction from circulating miRNA. Both the maximum depth and lambda parameter favored less model complexity and more L2 regularization (more conservative model). This may suggest mathematically that there may not be complex interactions between miRNA at least when predicting asthma exacerbations. One potential weakness of this preliminary analysis is that a small number of iterations were performed during randomized hyperparameter optimization as an extensive grid search would be computationally expensive.

A notable and important feature common to the logistic regression, neural network, and XGBoost models was that all analyses included hsa-miR-206 as a key feature as determined by both a univariate significance screen (parametric) and Kruskal Wallis

Test for binary and multiclass classification (non-parametric) regardless of median imputation or dense matrix analyses. The possible mechanistic significance of miR-206 is explored in the Discussion section of Chapter 4 and merits further biological exploration related to asthma.

In summary, resilient backpropagation and backpropagation neural networks did not yield good prediction at least in this miRNA dataset with the above limitations described. XGBoost did produce an AUROC similar to logistic regression although with different features (except miR-206) and usage of a dense matrix rather than median imputation. Additionally, the models do not seem to support potential non-linear complexity especially given the trend toward low model complexity and increased regularization in XGBoost. However, a common miRNA feature between all analyses is miR-206, which suggests robustness of this miRNA as a good predictor of asthma exacerbation.

Chapter 6 – Conclusion and Future Directions

This thesis explored circulating miRNA obtained at baseline prior to randomization to the inhaled corticosteroid arm in a cohort of childhood asthmatics and both the association with a measure of airways hyperresponsiveness (methacholine PC_{20}) and predictive capability of miRNA for asthma exacerbation with logistic regression, neural network, and XGBoost. First, the association analysis did reveal specific miRNA that mediated airways hyperresponsiveness with subsequent functional validation in human airway smooth muscle cells performed by collaborators. Subsequent analyses with logistic regression and machine learning techniques were performed to predict asthma exacerbation status in the year after randomization. A panel of 3 circulating miRNA alone and combined with a clinical asthma exacerbation score provided good prediction suggesting that the combination of 'omics with clinical information is important. Additional machine learning techniques did not exceed the predictive power of traditional logistic regression although XGBoost did produce an AUROC near the logistic regression AUROC for the logistic regression miRNA model. The main limitation of these preliminary machine learning analyses is that randomized hyperparameter optimization was performed in lieu of a grid search given computational complexity. Despite different feature selection techniques and data pre-processing steps, all techniques identified miR-206 as a key feature for prediction of exacerbation suggesting robust prediction with this particular miRNA.

Overall, ongoing study of miRNA in respiratory diseases such as asthma are merited and could yield novel biomarkers, pharmacogenomic insight, and possibly foster the development of new therapeutic drug classes[77]. Although logistic regression had the best performance for prediction of asthma exacerbation in this thesis, machine learning and analyses with greater sophistication are likely to supplant these traditional techniques. For instance, the cancer literature has utilized various supervised machine learning techniques and classification algorithms for prediction of disease outcomes with the postulation that "integration of multidimensional heterogeneous data, combined with the application of different techniques for feature selection and classification can

provide promising tools for inference.[78]" Integration of miRNA with other genomic and epigenomic data will also be necessary to gain a full understanding of disease states such as asthma. A model-based integration approach that "can merge predictive models from different data types" [79] may be one technique.

Appendices

SUPPLEMENTAL METHODS AND TABLES FOR CHAPTER 2

Supplemental Methods

A. CAMP Inclusion criteria
Inclusion criteria for CAMP were notable for children aged 5-12 years old at time of screening, chronic asthma symptoms for at least 6 months in the year prior to interview, PC_{20} < 12.5 mg/mL, and other factors. Children were excluded if it was determined that their asthma was severe, that there was a confounding or complicating condition, or that the child could not perform acceptable spirometry or methacholine challenge. Diary card data and other clinical characteristics were collected prior to randomization

B. CAMP Definition of asthma exacerbation and treatment protocol
Clinics instructed patients/caregivers to recognize an asthma exacerbation based on symptoms or by decrease in peak flow to < 80 % personal best. Oral prednisone was prescribed "if the patient uses more than 12 puffs of albuterol in 24 hours" or had symptom code described as "one or more asthma episodes that last longer than 2 hours or result in shortened normal activity, seeing a doctor for acute care, or going to a hospital for acute care" or if peak flow dropped to < 50 % of personal best despite bronchodilator usage. Prednisone dose and duration in addition to tapering protocol were specified in the protocol, and the physician could decide on an extended course of oral steroid depending on the clinical response [19].

C. Sample analysis excluded replicates
For replicate samples, only replicate set # 2 was included for analysis to minimize a subject contributing more than once to the data.

Table E1: Asthma Exacerbation Clinical Score

Reported asthma symptoms

1. Symptoms for >= 3 months of the year (1 point)
   Question 81 on Baseline Asthma and Allergy History. If symptoms were present for >= 3 months, one point awarded.
2. Symptoms precipitated by colds, cold air, exercise, dust (1 point each)
   Each Variable (Question 43, 46, 39, and 36, respectively) was coded from 0-3 with 0 meaning "never causes asthma" to 3 "always or almost always causes asthma. One point was awarded if the answer was 1,2, or 3.

Current Asthma Medications

1. Short-acting B2 agonist (1 point each)
   Question 24: Variable was coded from 1-5 with 4-5 meaning less than once per week and never, respectively. No point was awarded if scored as 4 or 5.
2. Inhaled steroids (1 point each)
   Question 20: In the last 6 months how often has your child used ICS for asthma? As with question 24, no point was awarded if scored as 4 or 5 (less than once per week and never, respectively)
3. Leukotriene inhibitors (1 point each)
   The CAMP trial did not use LTM.

Healthcare utilization

1. Ever hospitalized for asthma (1 point)
   Question 16: If the answer was (yes/1), 1 point awarded.
2. Ever admitted to ICU for asthma (1 point)
   Question 20: If the answer was (yes/1), 1 point awarded.
3. = 2 courses of steroids last year (1 point)
   Question 30: Referred to >= 2 courses of steroid in 6 months rather than 1 year. Most subjects did not require steroid therapy. If the answer was >=2, 1 point was awarded.
4. = 2 ER visits for asthma last year
   Question 21: If there was >=2 ER visits in a year, 1 point was awarded
5. Doctor visits last year - 3 or more (1 point)
   Question 22
6. Doctor visits last year - 6 or more (1 point)
   Question 22

Medical History

1. Personal history of eczema/hay fever (1 point)
   Questions 92 and 96
2. Parental history of asthma or atopy (1 point)
   Derived from parental history data – father and mother self-reported status was utilized

3. Smoke exposure as infant or current (1 point)
   Derived from parental history data – father and mother self-reported status was utilized

Total Score

1. Low risk (score ≤ 5)
2. Intermediate or average risk (score 6-8)
3. High risk (score ≥ 9)

SUPPLEMENTAL TABLES FOR CHAPTER 3

Table E2: Sensitivity Analysis for Circulatory miRNA Association by Least Squares Linear Regression with methacholine PC_{20} (univariate model, unranked) with outlier values removed and with detection of miRNA in at least 50 % of samples

miR	slope	p-value	FDR p-value	95 % CI Lower	95 % CI Upper
hsa-miR-296-5p	0.342	0.008	0.011	0.090	0.593
hsa-miR-548b-5p	0.353	0.003	0.069	0.127	0.578
hsa-miR-138-5p	0.361	0.005	0.090	0.109	0.613
hsa-miR-16-5p	0.238	0.010	0.109	0.059	0.417
hsa-miR-1227-3p	0.287	0.049	0.109	0.001	0.573
hsa-miR-30d-5p	0.202	0.084	0.109	-0.027	0.431
hsa-miR-203a-3p	0.042	0.042	0.120	0.008	0.388
hsa-miR-128-3p	0.642	0.009	0.187	0.162	1.123
hsa-miR-942-5p	0.232	0.033	0.211	0.019	0.446
hsa-miR-451a	0.228	0.017	0.211	0.042	0.414
hsa-miR-212-3p	0.454	0.080	0.240	-0.055	0.962
hsa-miR-143-3p	0.400	0.047	0.388	0.006	0.794
hsa-miR-638	0.148	0.353	0.474	-0.167	0.464
hsa-miR-25-3p	0.095	0.521	0.474	-0.196	0.385

Table E3: Circulatory miRNA Association by Least Squares Linear Regression with methacholine PC20 (multivariate model adjusting for age, sex, and height, unranked) with detection of miRNA in at least 50 % of samples

miR	miR slope	miR p-value	FDR p-value	95 % CI Lower	95 % CI Upper
hsa-miR-296-5p	0.501	0.00004	0.005	0.270	0.732
hsa-miR-548b-5p	0.340	0.002	0.080	0.127	0.554
hsa-miR-138-5p	0.380	0.003	0.081	0.136	0.624
hsa-miR-30d-5p	0.213	0.004	0.094	0.069	0.357
hsa-miR-16-5p	0.206	0.004	0.094	0.066	0.347
hsa-miR-1227-3p	0.323	0.006	0.120	0.094	0.552
hsa-miR-203a-3p	0.206	0.007	0.128	0.056	0.356
hsa-miR-128-3p	0.584	0.015	0.235	0.115	1.053
hsa-miR-451a	0.197	0.017	0.235	0.035	0.360
hsa-miR-942-5p	0.240	0.018	0.235	0.042	0.439
hsa-miR-212-3p	0.291	0.022	0.267	0.042	0.540
hsa-miR-143-3p	0.425	0.025	0.274	0.055	0.794
hsa-miR-185-5p	0.214	0.047	0.463	0.003	0.425
hsa-miR-25-3p	0.224	0.048	0.463	0.002	0.445

Table E4: Circulatory miRNA Association by Least Squares Linear Regression with methacholine PC20 (univariate model, ranked) with detection of miRNA in at least 50 % of samples

miR	miR slope	miR p-value	FDR p-value	95 % CI Lower	95 % CI Upper
hsa-miR-296-5p	0.307	0.0003	0.021	0.142	0.473
hsa-miR-138-5p	0.292	0.002	0.068	0.112	0.473
hsa-miR-16-5p	0.231	0.003	0.105	0.078	0.384
hsa-miR-451a	0.232	0.004	0.105	0.075	0.389
hsa-miR-548b-5p	0.263	0.009	0.151	0.068	0.457
hsa-miR-324-3p	0.285	0.009	0.151	0.072	0.499
hsa-miR-942-5p	0.213	0.009	0.151	0.053	0.373
hsa-miR-128-3p	0.276	0.010	0.151	0.068	0.483
hsa-miR-1227-3p	0.245	0.017	0.228	0.044	0.446
hsa-miR-92a-3p	0.192	0.018	0.228	0.034	0.351
hsa-let-7d-5p	0.202	0.028	0.330	0.023	0.381
hsa-miR-30d-5p	0.170	0.033	0.363	0.014	0.326
hsa-miR-145-5p	0.236	0.040	0.401	0.011	0.461
hsa-miR-181c-5p	0.205	0.042	0.401	0.007	0.404
hsa-miR-203a-3p	0.170	0.044	0.401	0.005	0.336

Table E5: Circulatory miRNA Association by Least Squares Linear Regression with methacholine PC20 (multivariate model adjusting for age, sex, and height, ranked) with detection of miRNA in at least 50 % of samples

miR	miR slope	miR p-value	FDR p-value	95 % CI Lower	95 % CI Upper
hsa-miR-296-5p	0.344	0.0001	0.016	0.173	0.515
hsa-miR-138-5p	0.305	0.001	0.054	0.121	0.490
hsa-miR-16-5p	0.243	0.003	0.082	0.086	0.400
hsa-miR-451a	0.235	0.004	0.104	0.076	0.393
hsa-miR-324-3p	0.330	0.005	0.110	0.103	0.557
hsa-miR-548b-5p	0.281	0.007	0.136	0.079	0.484
hsa-miR-942-5p	0.214	0.010	0.174	0.052	0.377
hsa-miR-128-3p	0.273	0.013	0.194	0.060	0.486
hsa-miR-92a-3p	0.199	0.016	0.232	0.037	0.360
hsa-miR-1227-3p	0.240	0.021	0.265	0.038	0.442
hsa-miR-30d-5p	0.181	0.025	0.296	0.023	0.338
hsa-let-7d-5p	0.198	0.033	0.364	0.016	0.379
hsa-miR-145-5p	0.245	0.036	0.369	0.017	0.473
hsa-miR-19b-3p	0.165	0.046	0.446	0.003	0.327

Table E6: miRBase Accession numbers for cytoscape (univariate model, unranked)

miR	miRBase Accession #
hsa-miR-296-5p	MIMAT0000690
hsa-miR-548b-5p	MIMAT0004798
hsa-miR-138-5p	MIMAT0000430
hsa-miR-16-5p	MIMAT0000069
hsa-miR-1227-3p	MIMAT0005580
hsa-miR-203a-3p	MIMAT0000264
hsa-miR-128-3p	MIMAT0000424
hsa-miR-30d-5p	MIMAT0000245

Table E7: DAVID Gene Ontology (GO) Analysis (GOTERM_BP_DIRECT)

Term	Number of Genes	Percentage of Genes (%)	P-value	Corrected P-value (Benjamini)
Translational initiation	57	2.1	$1.1x10^{-13}$	6.8×10^{-10}
rRNA processing	74	2.8	1.2×10^{-12}	3.6×10^{-9}
Protein stabilization	54	2.0	1.7×10^{-12}	3.5×10^{-9}
Positive regulation of transcription, DNA-templated	131	4.9	9.9×10^{-11}	1.5×10^{-7}
Negative regulation of apoptotic process	118	4.4	1.1×10^{-9}	1.3×10^{-6}

SUPPLEMENTAL TABLES AND FIGURES FOR CHAPTER 4

Table E8: Cross-validation (10-fold) of AUROC (all data)

	miR-clinical model	miR model	Clinical model
cvAUROC	0·74	0·66	0·66
95 % CI	0·66 – 0·82	0·57 – 0·75	0·56 – 0·75

Table E9: DAVID Biocarta Pathway Analysis

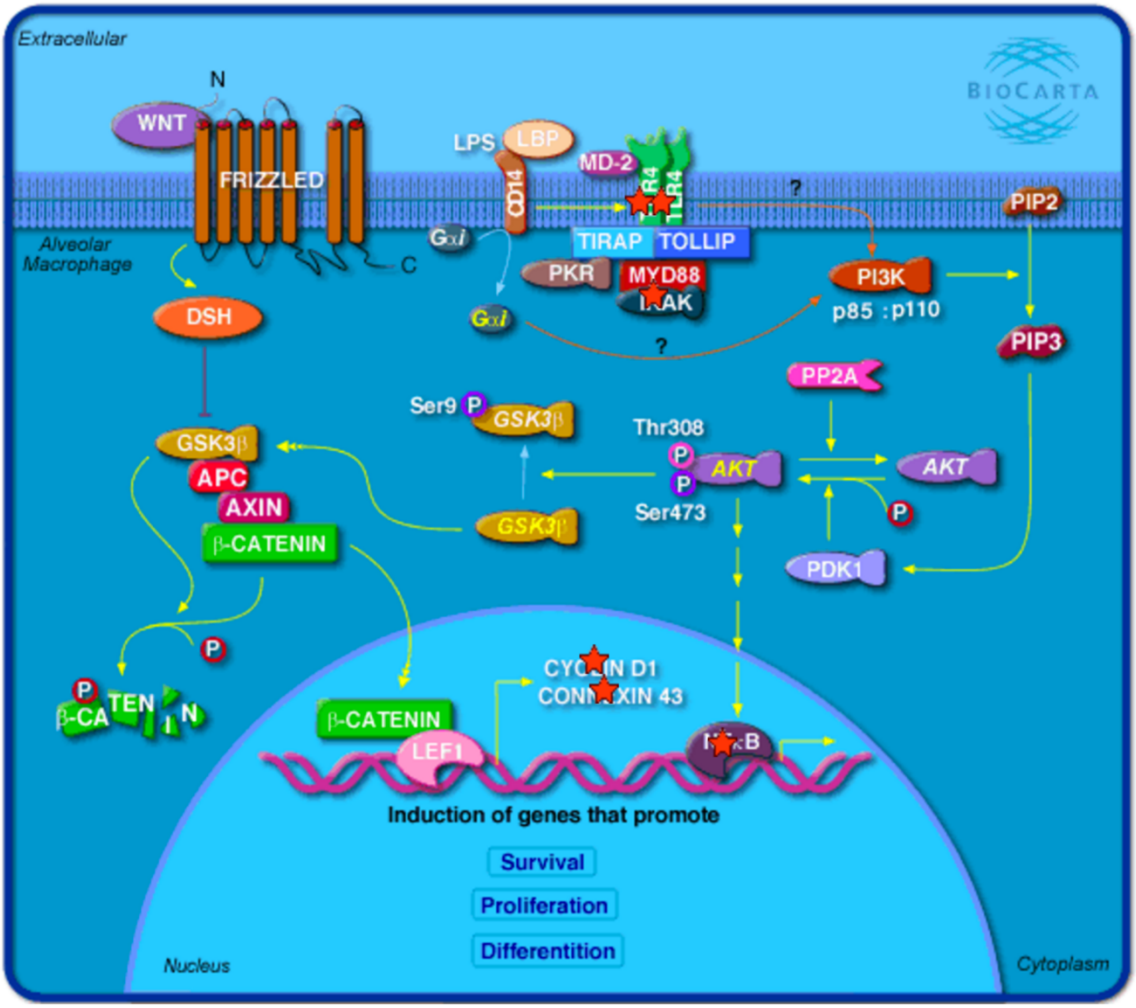
Pathway	Gene Count	P-value	Corrected P-value^
Inactivation of Gsk3 by AKT causes accumulation of b-catenin in Alveolar Macrophages	5	$2{\cdot}4x10^{-4}$	$1{\cdot}7x10^{-2}$
Influence of Ras and Rho proteins on G1 to S Transition	4	$3{\cdot}9x10^{-3}$	$6{\cdot}5x10^{-2}$
Cyclins and Cell Cycle Regulation	4	$3{\cdot}1x10^{-3}$	$8{\cdot}0x10^{-2}$
NF-κβ Signaling Pathway	4	$2{\cdot}3x10^{-3}$	$8{\cdot}0x10^{-2}$

^ Benjamini-Hochberg correction

Table E10: Top DAVID GOTERM_BP DIRECT

Term	Gene Count	Percentage of Genes	P-value	Corrected P-value^
Positive regulation of fibroblast proliferation	4	12·9	$1{\cdot}2x10^{-4}$	$1{\cdot}5x10^{-2}$
Positive regulation of phospholipase C activity	3	9·7	$1{\cdot}1x10^{-4}$	$1{\cdot}8x10^{-2}$
Pentose-phosphate shunt	3	9·7	$1{\cdot}7x10^{-4}$	$1{\cdot}8x10^{-2}$
Positive Regulation of gene expression	6	19·4	$9{\cdot}2x10^{-5}$	$2{\cdot}0x10^{-2}$
Negative regulation of transcription from RNA polymerase II promoter	8	25·8	$2{\cdot}2x10^{-4}$	$2{\cdot}0x10^{-2}$

^ Benjamini-Hochberg correction



Supplemental Figure 1: DAVID (Database from Annotation, Visualization, and Integrated Discovery) Biocarta Pathway analysis - Inactivation of Gsk3 by AKT causes accumulation of b-catenin in alveolar macrophages. miRTarBase was used to determine experimentally validated microRNA-target interactions with genes. The gene list was subsequently used for pathway analysis. The genes marked with the red star are targeted by the microRNA. Illustration by BioCarta (http://www.biocarta.com/), released under Creative Commons Attribution License – Attribution-NonCommercial 4.0 International (CC BY-NC 4.0), 2017. Permission is only for use of Biocarta's figure. Please refer to BioCarta's Disclaimer of Liability and of Warranties on its website for further information.

SUPPLEMENTAL TABLES AND FIGURES FOR CHAPTER 5

None

References

1. To T, Stanojevic S, Moores G, Gershon AS, Bateman ED, Cruz AA, et al. Global asthma prevalence in adults: findings from the cross-sectional world health survey. BMC public health. 2012;12:204.
2. Fanta CH. Asthma. The New England journal of medicine. 2009;360(10):1002-14.
3. Bahadori K, Doyle-Waters MM, Marra C, Lynd L, Alasaly K, Swiston J, et al. Economic burden of asthma: a systematic review. BMC pulmonary medicine. 2009;9:24.
4. Jang J, Gary Chan KC, Huang H, Sullivan SD. Trends in cost and outcomes among adult and pediatric patients with asthma: 2000-2009. Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology. 2013;111(6):516-22.
5. Robroeks CM, van Vliet D, Jobsis Q, Braekers R, Rijkers GT, Wodzig WK, et al. Prediction of asthma exacerbations in children: results of a one-year prospective study. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 2012;42(5):792-8.
6. Haselkorn T, Zeiger RS, Chipps BE, Mink DR, Szefler SJ, Simons FE, et al. Recent asthma exacerbations predict future exacerbations in children with severe or difficult-to-treat asthma. The Journal of allergy and clinical immunology. 2009;124(5):921-7.
7. Xu M, Tantisira KG, Wu A, Litonjua AA, Chu JH, Himes BE, et al. Genome Wide Association Study to predict severe asthma exacerbations in children using random forests classifiers. BMC medical genetics. 2011;12:90.
8. Weiland M, Gao XH, Zhou L, Mi QS. Small RNAs have a large impact: circulating microRNAs as biomarkers for human diseases. RNA biology. 2012;9(6):850-9.
9. Witwer KW. Circulating microRNA biomarker studies: pitfalls and potential solutions. Clinical chemistry. 2015;61(1):56-63.
10. Hoy AM, Buck AH. Extracellular small RNAs: what, where, why? Biochemical Society transactions. 2012;40(4):886-90.
11. Russo F, Di Bella S, Nigita G, Macca V, Lagana A, Giugno R, et al. miRandola: extracellular circulating microRNAs database. PloS one. 2012;7(10):e47786.
12. Sheinerman KS, Tsivinsky VG, Crawford F, Mullan MJ, Abdullah L, Umansky SR. Plasma microRNA biomarkers for detection of mild cognitive impairment. Aging. 2012;4(9):590-605.
13. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(30):10513-8.
14. Tijsen AJ, Pinto YM, Creemers EE. Circulating microRNAs as diagnostic biomarkers for cardiovascular diseases. American journal of physiology Heart and circulatory physiology. 2012;303(9):H1085-95.

15. Booton R, Lindsay MA. Emerging role of MicroRNAs and long noncoding RNAs in respiratory disease. Chest. 2014;146(1):193-204.
16. Sheinerman KS, Tsivinsky VG, Umansky SR. Analysis of organ-enriched microRNAs in plasma as an approach to development of Universal Screening Test: feasibility study. Journal of translational medicine. 2013;11:304.
17. Panganiban RP, Wang Y, Howrylak J, Chinchilli VM, Craig TJ, August A, et al. Circulating microRNAs as biomarkers in patients with allergic rhinitis and asthma. The Journal of allergy and clinical immunology. 2016;137(5):1423-32.
18. Recruitment of participants in the childhood Asthma Management Program (CAMP). I. Description of methods: Childhood Asthma Management Program Research Group. The Journal of asthma : official journal of the Association for the Care of Asthma. 1999;36(3):217-37.
19. The Childhood Asthma Management Program (CAMP): design, rationale, and methods. Childhood Asthma Management Program Research Group. Control Clin Trials. 1999;20(1):91-120.
20. Weiss ST, Van Natta ML, Zeiger RS. Relationship between increased airway responsiveness and asthma severity in the childhood asthma management program. American journal of respiratory and critical care medicine. 2000;162(1):50-6.
21. Kho AT, Sharma S, Davis JS, Spina J, Howard D, McEnroy K, et al. Circulating MicroRNAs: Association with Lung Function in Asthma. PloS one. 2016;11(6):e0157998.
22. Forno E, Fuhlbrigge A, Soto-Quiros ME, Avila L, Raby BA, Brehm J, et al. Risk factors and predictive clinical scores for asthma exacerbations in childhood. Chest. 2010;138(5):1156-65.
23. Zhu W, Kan X. Neural network cascade optimizes microRNA biomarker selection for nasopharyngeal cancer prognosis. PloS one. 2014;9(10):e110537.
24. Lai X, Wolkenhauer O, Vera J. Understanding microRNA-mediated gene regulatory networks through mathematical modelling. Nucleic acids research. 2016;44(13):6019-35.
25. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nature reviews Genetics. 2015;16(6):321-32.
26. Chen T, Guestrin,C. XGBoost: A Scalable Tree Boosting System. KDD '16 Proceeding of the 22nd ACM SIGKDD International Conference on Knowledge and Discovery and Data Mining 2016:785-94.
27. Wang Y, Yang L, Li P, Huang H, Liu T, He H, et al. Circulating microRNA Signatures Associated with Childhood Asthma. Clinical laboratory. 2015;61(5-6):467-74.
28. Jardim MJ, Dailey L, Silbajoris R, Diaz-Sanchez D. Distinct microRNA expression in human airway cells of asthmatic donors identifies a novel asthma-associated gene. American journal of respiratory cell and molecular biology. 2012;47(4):536-42.
29. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic acids research. 2008;36(Database issue):D154-8.
30. Huang RS, Gamazon ER, Ziliak D, Wen Y, Im HK, Zhang W, et al. Population differences in microRNA expression and biological implications. RNA biology. 2011;8(4):692-701.
31. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015.

32. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498-504.
33. Kutmon M, Kelder T, Mandaviya P, Evelo CT, Coort SL. CyTargetLinker: a cytoscape app to integrate regulatory interactions in network analysis. PloS one. 2013;8(12):e82160.
34. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic acids research. 2011;39(Database issue):D163-9.
35. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research. 2009;37(1):1-13.
36. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic acids research. 2016;44(D1):D457-62.
37. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28(1):27-30.
38. Hu R, Pan W, Fedulov AV, Jester W, Jones MR, Weiss ST, et al. MicroRNA-10a controls airway smooth muscle cell proliferation via direct targeting of the PI3 kinase pathway. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2014;28(5):2347-57.
39. National Asthma E, Prevention P. Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma-Summary Report 2007. The Journal of allergy and clinical immunology. 2007;120(5 Suppl):S94-138.
40. Subhash R. Lele JLK, Peter Solymos. ResourceSelection: Resource Selection (Probability) Functions for Use-Availability
Data, R package version 0.2-6. 2016.
41. Achim Zeileis TH. Diagnostic Checking in Regression Relationships. R News. 2002;2(3):7-10.
42. Kuhn MCfJW, Steve Weston, Andre Williams, Chris, Keefer AE, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R, Core Team MB, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca,, Candan YTaC. caret: Classification and Regression Training. R package version 6.0-64. 2016.
43. Xavier Robin NT, Alexandre Hainard,Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, Markus Müller. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
44. T. Sing OS, N. Beerenwinkel, T. Lengauer. ROCR: visualizing classifier performance in R. Bioinformatics. 2005;21(20):7881.
45. Erin LeDell MP, Mark van der Laan. cvAUC: Cross-Validated Area Under the ROC Curve Confidence Intervals, R package version 1.1.0. 2014.
46. Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic acids research. 2016;44(D1):D239-47.
47. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44-57.

48. Bischl B, et al. mlr: Machine Learning in R. Journal of Machine Learning Research. 2016;17.
49. Chung S, Lee TJ, Reader BF, Kim JY, Lee YG, Park GY, et al. FoxO1 regulates allergic asthmatic inflammation through regulating polarization of the macrophage inflammatory phenotype. Oncotarget. 2016;7(14):17532-46.
50. Zhou J, Xu F, Yu JJ, Zhang W. YAP is up-regulated in the bronchial airway smooth muscle of the chronic asthma mouse model. International journal of clinical and experimental pathology. 2015;8(9):11132-9.
51. Robson JE, Eaton SA, Underhill P, Williams D, Peters J. MicroRNAs 296 and 298 are imprinted and part of the GNAS/Gnas cluster and miR-296 targets IKBKE and Tmed9. Rna. 2012;18(1):135-44.
52. Patel MN, Bernard WG, Milev NB, Cawthorn WP, Figg N, Hart D, et al. Hematopoietic IKBKE limits the chronicity of inflammasome priming and metaflammation. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(2):506-11.
53. Edwards MR, Bartlett NW, Clarke D, Birrell M, Belvisi M, Johnston SL. Targeting the NF-kappaB pathway in asthma and chronic obstructive pulmonary disease. Pharmacology & therapeutics. 2009;121(1):1-13.
54. Sheller JR, Polosukhin VV, Mitchell D, Cheng DS, Peebles RS, Blackwell TS. Nuclear factor kappa B induction in airway epithelium increases lung inflammation in allergen-challenged mice. Experimental lung research. 2009;35(10):883-95.
55. Ziegelbauer K, Gantner F, Lukacs NW, Berlin A, Fuchikami K, Niki T, et al. A selective novel low-molecular-weight inhibitor of IkappaB kinase-beta (IKK-beta) prevents pulmonary inflammation and shows broad anti-inflammatory activity. British journal of pharmacology. 2005;145(2):178-92.
56. Solberg OD, Ostrin EJ, Love MI, Peng JC, Bhakta NR, Hou L, et al. Airway epithelial miRNA expression is altered in asthma. American journal of respiratory and critical care medicine. 2012;186(10):965-74.
57. Williams AE, Larner-Svensson H, Perry MM, Campbell GA, Herrick SE, Adcock IM, et al. MicroRNA expression profiling in mild asthmatic human airways and effect of corticosteroid therapy. PloS one. 2009;4(6):e5889.
58. Moulton DE, Goriely A. Possible role of differential growth in airway wall remodeling in asthma. Journal of applied physiology. 2011;110(4):1003-12.
59. Lv Y, Qi R, Xu J, Di Z, Zheng H, Huo W, et al. Profiling of serum and urinary microRNAs in children with atopic dermatitis. PloS one. 2014;9(12):e115448.
60. Ke XF, Fang J, Wu XN, Yu CH. MicroRNA-203 accelerates apoptosis in LPS-stimulated alveolar epithelial cells by targeting PIK3CA. Biochemical and biophysical research communications. 2014;450(4):1297-303.
61. Dekkers BG, Maarsingh H, Meurs H, Gosens R. Airway structural components drive airway smooth muscle remodeling in asthma. Proceedings of the American Thoracic Society. 2009;6(8):683-92.
62. Martin JG, Ramos-Barbon D. Airway smooth muscle growth from the perspective of animal models. Respir Physiol Neurobiol. 2003;137(2-3):251-61.
63. Grasedieck S, Scholer N, Bommer M, Niess JH, Tumani H, Rouhi A, et al. Impact of serum storage conditions on microRNA stability. Leukemia. 2012;26(11):2414-6.

64. Mattes J, Collison A, Plank M, Phipps S, Foster PS. Antagonism of microRNA-126 suppresses the effector function of TH2 cells and the development of allergic airways disease. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(44):18704-9.
65. Collison A, Mattes J, Plank M, Foster PS. Inhibition of house dust mite-induced allergic airways disease by antagonism of microRNA-145 is comparable to glucocorticoid treatment. The Journal of allergy and clinical immunology. 2011;128(1):160-7 e4.
66. Kumar M, Ahmad T, Sharma A, Mabalirajan U, Kulshreshtha A, Agrawal A, et al. Let-7 microRNA-mediated regulation of IL-13 and allergic airway inflammation. The Journal of allergy and clinical immunology. 2011;128(5):1077-85 e1-10.
67. Chalise SP, Bhatta NK, Singh RR, Prasad MS, Poudel P. Assessment of control of bronchial asthma in children using Childhood Asthma Control Test. The Indian journal of chest diseases & allied sciences. 2014;56(2):75-8.
68. Comer BS, Camoretti-Mercado B, Kogut PC, Halayko AJ, Solway J, Gerthoffer WT. MicroRNA-146a and microRNA-146b expression and anti-inflammatory function in human airway smooth muscle. American journal of physiology Lung cellular and molecular physiology. 2014;307(9):L727-34.
69. Garbacki N, Di Valentin E, Huynh-Thu VA, Geurts P, Irrthum A, Crahay C, et al. MicroRNAs profiling in murine models of acute and chronic asthma: a relationship with mRNAs targets. PloS one. 2011;6(1):e16509.
70. Lu TX, Rothenberg ME. Diagnostic, functional, and therapeutic roles of microRNA in allergic diseases. The Journal of allergy and clinical immunology. 2013;132(1):3-13; quiz 4.
71. Collison A, Siegle JS, Hansbro NG, Kwok CT, Herbert C, Mattes J, et al. Epigenetic changes associated with disease progression in a mouse model of childhood allergic asthma. Disease models & mechanisms. 2013;6(4):993-1000.
72. Clifford RL, Singer CA, John AE. Epigenetics and miRNA emerge as key regulators of smooth muscle cell phenotype and function. Pulmonary pharmacology & therapeutics. 2013;26(1):75-85.
73. Bentley JK, Deng H, Linn MJ, Lei J, Dokshin GA, Fingar DC, et al. Airway smooth muscle hyperplasia and hypertrophy correlate with glycogen synthase kinase-3(beta) phosphorylation in a mouse model of asthma. American journal of physiology Lung cellular and molecular physiology. 2009;296(2):L176-84.
74. Ge Q, Zhou Y, Lu J, Bai Y, Xie X, Lu Z. miRNA in plasma exosome is stable under different storage conditions. Molecules. 2014;19(2):1568-75.
75. Shaughnessy RG, Farrell D, Riepema K, Bakker D, Gordon SV. Analysis of Biobanked Serum from a Mycobacterium avium subsp paratuberculosis Bovine Infection Model Confirms the Remarkable Stability of Circulating miRNA Profiles and Defines a Bovine Serum miRNA Repertoire. PloS one. 2015;10(12):e0145089.
76. Olden J, Joy, M, Death, R. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecological Modelling. 2004;178(3-4):389-97.
77. Santulli G. Exploiting microRNA Specificity and Selectivity: Paving a Sustainable Path Towards Precision Medicine. Advances in experimental medicine and biology. 2015;888:1-3.

78. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal. 2015;13:8-17.
79. Lin E, Lane HY. Machine learning and systems genomics approaches for multi-omics data. Biomarker research. 2017;5:2.