

Interpretation of Hemes: A Human Metabolomic Serum Biomarker of Insufficient Sleep

Paul Bisesi

University of Colorado Boulder
Department of Molecular, Cellular, and Developmental Biology
Undergraduate Honors Thesis
April 4th, 2018

Committee Members

Dr. Kenneth P. Wright Jr. (IPHY), Thesis Advisor
Dr. Pamala Harvey (MCDB), Honors Council Representative
Dr. Bradley Olwin (MCDB)

Table of Contents

Abstract	01
Introduction	02
Methods	08
<i>Participants</i>	08
<i>Study Protocol</i>	09
<i>Blood Collection and Processing</i>	11
<i>Data Preparation</i>	11
<i>Data Analysis</i>	12
<i>Independent Protocol Participants</i>	13
<i>Independent Protocol Procedure</i>	13
<i>Independent Protocol Blood Collection and Processing</i>	15
<i>Independent Protocol Data Preparation</i>	15
<i>Independent Protocol Data Analysis</i>	15
Results	16
<i>Biomarker Creation and Score</i>	16
<i>Biomarker Score AUC</i>	16
<i>Biomarker Score 50% Holdout Test</i>	17
<i>Independent Protocol Biomarker Creation and Score</i>	18
<i>Independent Protocol Biomarker Score AUC</i>	19
<i>Independent Protocol Biomarker Score 50% Holdout Test</i>	20
Discussion	21
Acknowledgements	26
References	27

Abstract

Advances in computational power and molecular detection techniques allow investigators unparalleled insight into the metabolome, the collection of small-molecule metabolism products, called metabolites, found in organisms. Metabolomics, the study of such molecules, attempts to discern how various biological conditions such as disease or stress alters metabolite concentration in the body. Investigators can also leverage metabolomics to create biomarkers, tools used to monitor and diagnose biological conditions. Metabolomic biomarkers are often created using metabolites found in accessible biofluids, like blood or urine, and are usually developed to diagnose a disease. Insufficient sleep impacts over a third of Americans, and results in metabolic changes, making it a notable target for biomarker development. The current study uses a cohort of 16 human participants experiencing a protocol of sufficient and insufficient sleep to develop a biomarker using blood metabolite concentration data. Principal component analysis, partial least squares discriminant analysis, and area under the receiver operating characteristic curve analysis aided in the creation of a final biomarker score consisting of six metabolites, weighted by their contribution to overall performance. Biomarker performance was assessed by predictive accuracy at classifying samples into sufficient and insufficient sleep conditions, and area under the receiver operating characteristic curve analysis. These tests were performed in both the original dataset and an independent dataset consisting of 36 individuals experiencing a different protocol of insufficient sleep. Independent verification is essential for effective biomarker development to ensure the results are applicable to a wide range of individuals.

Introduction

Discerning the underlying biological processes occurring in humans creates many challenges. How can a physician obtain an objective assessment of a patient's health, or determine the accuracy of a self-reported health history? Are a patient's symptoms induced by an underlying condition, or are they benign? Biomarkers are one tool capable of elucidating these questions. The National Institutes of Health defines a biomarker as a "structure, substance, or process that can be measured in the body" which can predict an "outcome or disease"¹. Current techniques utilizing biomarkers include breathalyzer tests, which determine blood alcohol content by measuring ethanol oxidation from an individual's breath, and saliva assays measuring melatonin concentration, which can assess circadian phase². Using biomarkers, investigators can obtain critical evidence about an individual's biological state, allowing them to analyze health problems escaping easy diagnoses, such as sleep difficulties.

The Centers for Disease Control and Prevention report over a third of individuals in the United States obtain less than the recommended 7 hours of sleep per night^{3,4}. Sleeping below this amount can contribute to a variety of negative outcomes, including an increased risk for diabetes, cancer, and anxiety, and decreased performance in various tasks, including driving and academics⁵⁻⁹. Roughly one-fourth of Americans suffer from a sleep disorder, including sleep apnea or insomnia, suggesting sleep disorders contribute to the insufficient sleep epidemic¹⁰. Studies suggest sleep disorders are also underdiagnosed in the populace, indicating a need for increased awareness and diagnostic technique availability^{11,12}. However, current techniques for diagnosing sleep difficulties often require patients undergo expensive, time-consuming sleep evaluation visits at a sleep clinic, precluding those with limited funds or time from undergoing these procedures. Other, cheaper diagnostic techniques, such as sleep diaries or questionnaires,

can misrepresent the total amount of sleep an individual receives, potentially causing inaccurate diagnoses^{13,14}. A biomarker sensitive to insufficient sleep may allow inexpensive, accurate determination of an individual's sleep health, potentially mitigating these problems found with other diagnostic techniques. If a biomarker diagnoses an individual with insufficient sleep, their physician could then prescribe additional tests for sleep disorders, saving resources by preventing unneeded testing.

Metabolomics, the study of small-molecule metabolism products called metabolites, provides a powerful tool for biomarker creation¹⁵. Using these techniques allows investigators to analyze physiological changes, and use observed variation to develop a biomarker. Blood metabolites exhibit circadian rhythmicity in both human and animal models, making the blood an attractive target for insufficient sleep biomarker development¹⁶. A blood-based biomarker also provides straightforward sample collection; determining an individual's sleep health could be accomplished as part of a routine blood draw during a physical. In mice, investigators used metabolomic techniques to develop a blood-based biomarker of sleep restriction, suggesting the feasibility of creating a similar marker in humans¹⁷. When compared to other commonly used biofluids in metabolomic development, blood theoretically reflects metabolism occurring in all tissues, while a biofluid like urine contains metabolites from a more limited subset of tissue¹⁸. Insufficient sleep impacts a range of metabolic pathways, including glucose metabolism¹⁹. Thus, the wide range of metabolites found within blood, and sensitivity of the fluid to circadian effects, makes it an ideal biofluid to test for the physiological effects of insufficient sleep²⁰.

Once collected, blood samples must be processed to quantify their metabolite concentration. In metabolomics research, two major strategies exist for this step: targeted and untargeted strategies. Researchers using targeted metabolomics select a specific cohort of

metabolites to analyze in a sample²¹. Such an approach requires a full understanding of the biological system in question, to ensure the chosen metabolites are biologically relevant, but grants a clear picture of changes in those specific metabolites. Conversely, an untargeted metabolomic approach attempts to quantify all detectable metabolites found in a given sample²². This approach is beneficial for discovery focused analyses, where the metabolites of interest are unknown, as it theoretically captures the detectable range of metabolite changes, which ideally comprises most of the variation. Untargeted metabolomics also detects novel metabolites, those molecules not yet categorized in a database, which can create further opportunities for investigation²². The unclear metabolic changes caused by insufficient sleep lends itself well to an untargeted metabolomics approach to biomarker development.

Researchers use a variety of platforms to quantify metabolite concentration in a sample; mass spectrometry is one of the more common methods used for biomarker development. This technology ionizes the molecules found within a sample, separates them based on their mass-to-charge ratios, and feeds them into a detector which records the relative intensity of each detected metabolite, reported as the height of the observed intensity peak with arbitrary units²³. The mass spectrometer also records the retention time for each metabolite, which represents the amount of time a metabolite takes to pass through the detector²⁴. These samples are often fractionated into lipid and aqueous phases prior to this procedure to maximize the identification of the different types of metabolites. Cross referencing the mass spectrometer output with several online metabolite databases, which catalog the peak heights and retention times of various metabolites, allows identification of many of the specific metabolites found in a sample^{25,26}.

A variety of software packages and tools exist to assist in processing and interpreting metabolomics data. One of these tools, a web server called MetaboAnalyst, contains a range of

statistical procedures commonly performed in metabolomics analysis²⁷. Two major divisions exist in metabolomic statistics: exploratory and functional analysis. Exploratory analysis refers to a researcher's investigation into the statistically relevant metabolites present in a given sample to find a relationship with a disease condition, while functional analysis refers to a researcher investigating known, biologically relevant metabolites to determine their physiological significance. Developing a sleep biomarker begins by employing exploratory metabolomics. Multivariate analysis, especially principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), are two popular measures to explore metabolomic data^{28,29}. PCA takes a set of multi-dimensional data, such as a table of metabolite concentrations, and reduces its structure to a set of principal components, each which attempts to describe as much variability found in the data as possible³⁰. PLS-DA acts as a supervised version of PCA, meaning the algorithm knows which condition a given sample belongs to before performing its analysis³¹. Utilizing PCA and PLS-DA techniques allows investigators to determine the general data structure of their metabolomics data, providing opportunities to choose metabolites of interest or address any structural concerns caused by sample preparation.

Structural data trends arising from sample preparation may require correction before proceeding with analysis. Processing the large amounts of data generated in metabolomics using mass spectroscopy can introduce variations in the data caused solely by external events, not by any biological relevance found in the data itself. For example, conditions such as atmospheric ozone levels, or performing mass spectrometry on different days, can introduce these changes^{32,33}. These variations are called batch effects, where a batch refers to a set of samples measured together. Empirical Bayes methods, including the ComBat (Combating Batch Effects When Combining Batches of Gene Expression Data) method, are a common solution for

correcting batch effects. ComBat estimates the data's previous distribution to correct it, and especially works well for correcting batch errors in data containing numerous, small batches, those with fewer than 25 samples^{34,35}. Figure 1 shows an example of batch effects before and after ComBat batch correction, visualized using PCA. Attempting biomarker development on uncorrected batch data can result in investigators inappropriately assigning biomarkers based on technical errors instead of biological relevance, underlining the importance of batch correction^{36,37}.

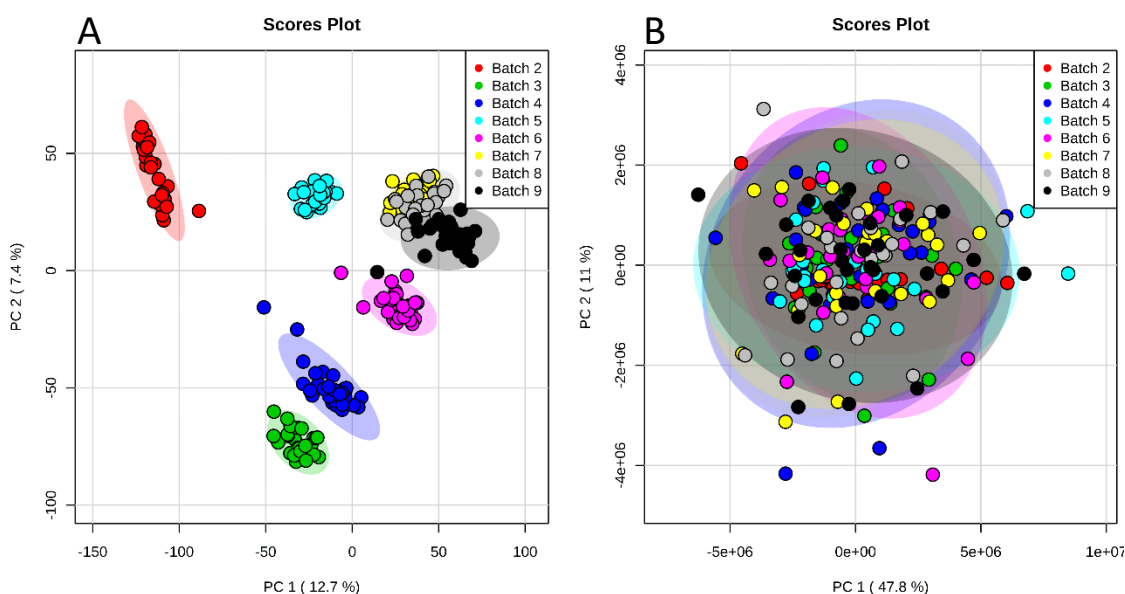


Figure 1. Batch Effect Correction in the Aqueous Phase. (A) Principal component analysis displays clear batch effects in the raw mass spectrometer data, where each batch reflects samples run concurrently. (B) Batch correction using the ComBat algorithm mitigates these effects.

Missing data comprise another common occurrence while utilizing mass spectroscopy in metabolomics procedures. A missing value may be caused by biological reasons, such as the metabolite concentration in a specific sample laying below the detection threshold, or technical problems³⁸. Studies suggest missing data in mass spectrometry experiments typically affects up to 20% of data, indicating the phenomenon is common^{39,40}. Several methods exist to handle

missing data, including ignoring variables containing missing data and estimating missing data using imputation techniques^{41,42}. Since most metabolomic data contains a notable fraction of missing data, the first strategy only applies to studies with small metabolite pools. Many other types of analysis, including PCA and PLS-DA, also require complete data sets and would be underpowered without the additional data provided by imputation⁴³. However, imputation does not come without cost, as applying the method inappropriately can lead to incorrect results⁴⁴. Bayesian Principal Component Analysis (BPCA) is a method incorporating aspects of PCA and Bayesian statistics to impute values based on prior knowledge about the data's distribution⁴⁵. Studies suggest BPCA performs better than other imputation methods, such as singular value decomposition imputation, based on the algorithm's performance predicting a set of artificially removed missing values^{46,47}.

Once batch correction and imputation methods are performed on the data, the search for a biomarker can begin in earnest. Several methods exist to determine significant metabolites warranting further investigation. One of these methods, the receiver operator characteristic (ROC) curve, displays the probability of a classifier correctly classifying a randomly chosen sample into its proper condition^{48,49}. In terms of testing a sleep biomarker, a high performing biomarker would accurately classify samples into either sufficient or insufficient sleep conditions. Such ROC analyses are considered the standard for clinical applications in other fields, and are gaining acceptance in the metabolomics biomarker literature^{15,50-52}. An ROC curve is plotted comparing the true positive rate, also known as the sensitivity, against the false positive rate, known as the specificity, with values from 0.0 to 1.0. The metric is often reported as the area under the curve (AUC), with a perfect biomarker having an AUC of 1.0, and one which predicts no better than chance receiving a score of 0.5. Kleinbaum and Klein established

cut-offs for the discrimination achieved by an AUC as follows: 0.5-0.6 = failed, 0.6-0.7 = poor, 0.7-0.8 = fair, 0.8-0.9 = good, and 0.9-1.0 = excellent⁵³. Investigators can employ ROC analysis to not only choose metabolites of interest to incorporate into a biomarker, but also to test the performance of the final biomarker at discriminating sample assignments.

Bennett and Devarajan describe characteristics composing an ideal biomarker, including: high specificity for the disease condition, high sensitivity, easy accessibility in blood or urine, robust prognostic value for outcomes, and providing insight into a disease mechanism⁵⁴. A biomarker composed of multiple metabolites typically fulfills more of these requirements, especially the sensitivity and specificity criteria, than do biomarkers composed of a single metabolite¹⁵. A multiple metabolite biomarker typically takes the form of a biomarker score, which weights metabolite concentrations by their contribution prediction performance, and sums them together to create a final score^{55,56}. Such a score is often created using β -coefficients from binomial logistic regression analysis as the weighting factor⁵⁷. External verification of any discovered biomarker in an independent dataset also comprises an essential factor in creating a robust biomarker, as a biomarker holding poor predictive value in a wide range of datasets holds little value in a clinical setting⁵⁸. This study attempts to create a biomarker for insufficient sleep addressing each of the listed criteria for an ideal biomarker, and validate it in an independent dataset.

Methods

Participants

We studied sixteen participants (eight women) ages 22.4 ± 4.8 years, all with a self-reported typical sleep schedule of 8.26 ± 0.69 hours. The University of Colorado Institutional Review Board, the Colorado Multiple Institutional Review Board, and the Colorado Clinical and

Translational Sciences Institute's Scientific Advisory and Review Committee approved all protocols, and all participants gave informed written consent prior to study. Using a variety of health screening procedures, including a physical examination, clinical psychiatric interview, electrocardiogram, metabolic panel, toxicology screen, and polysomnographic sleep disorders assessment, we determined the participants free of medical and psychological disorders.

Inclusion criteria for the study included: ages 18-35 years old, a habitual sleep schedule between 7 and 9.25 hours of sleep per night, less than 500 mg/day caffeine use, non-smokers, a BMI between 18.5 and 24.9 kg/m², and no drug dependence. Exclusion criteria included: medical and psychiatric conditions, shift work, pregnancy, lifetime BMI over 27.5 kg/m², living below an altitude of 1,600 meters during the prior year, traveling more than one time zone 3 weeks before the study, or recent weight loss. We selected participants who typically performed low levels of physical activity to control for the low levels of activity experienced during study visits.

Participants completed a urine toxicology test at the health screening appointment and prior to the study to confirm drug-free status.

Study Protocol

Participants stopped caffeine consumption and maintained a consistent, 9-hour sleep schedule for one week prior to the study. Wrist actigraphy, daily call-ins of bed and wake times, and sleep diary completion ensured adherence to assigned sleep schedules. Three days prior to the study, participants received meals composed of 30% fat, 55% carbohydrates, and 15% protein, reflecting the average US diet. The Clinical Translational Nutrition Core prepared the meals using participants' resting metabolic rate data, collected during the health screening, to ensure participants maintained their energy balance prior to undertaking the study. We

additionally proscribed exercise and beverages besides water to ensure energy balance. The inpatient protocol began with a 3-day baseline assessment, consisting of 9-hour sleep opportunities scheduled at each participant's habitual bedtime based on pre-study monitoring. The remainder of the study consisted of two, 5-day long sequences: a 9-hour condition followed by a sleep restriction condition, or the sleep restriction condition followed by the 9-hour condition. We assigned participants to one of the two conditions at random. The 9-hour condition gave participants of 9-hour sleep opportunities each night of that condition, while the sleep restriction condition gave them 5-hour sleep opportunities each night (Figure 2). During the baseline days, participants consumed the same energy balanced diet they received before the study. On study days, participants received meals containing 130-150% more calories than diets given during baseline days, of which they could eat as much as they wanted, and *ad libitum* snacks. During the

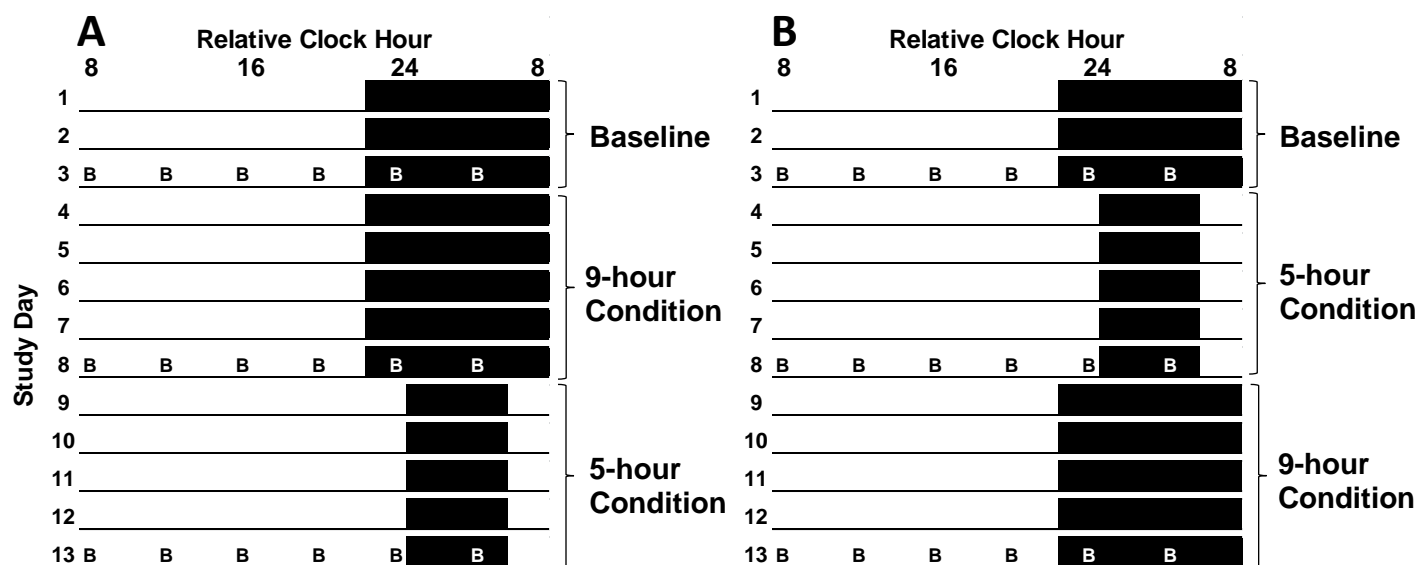


Figure 2. Biomarker Development Study Protocol. Participants completed two experimental conditions following their baseline assessment: a 9-hour condition and a 5-hour condition. Participants received a 9-hour sleep opportunity each night during the 9-hour condition and a 5-hour sleep opportunity during the 5-hour condition. We randomly assigned participants to one of two treatment groups: the 9-hour condition followed by the 5-hour condition (A) or the 5-hour condition followed by the 9-hour condition (B). Blood samples were taken every 4 hours during the final 24 hours of each condition, signified by a “B,” and sleep opportunities are labeled in black.

study, participants completed twice-daily stepping exercises to mimic the low levels of activity performed outside of the laboratory environment.

Blood Collection and Processing

We drew participants' blood every 4 hours during the last 24 hours of each of the three conditions. The first sample of each day was a fasted sample, drawn before the participant ate. Samples were centrifuged and stored at -80 °C until analysis. These samples were then processed for mass spectroscopy. Proteins were precipitated with methanol and extracted using MTBE. Aqueous molecules were separated using hydrophilic interaction liquid chromatography, and lipid molecules were separated using C18. Each fraction was then analyzed using a QTOF mass spectrometer (Agilent). Initial metabolite identification was performed using the Metlin, Human Metabolome Database, Lipid Maps, and Kyoto Encyclopedia of Genes and Genomes databases^{25,26,59,60}.

Data Preparation

Lipid and aqueous mass spectrometer peak height analysis data files was filtered to remove metabolites with missing data in found in 50% or greater of samples. These data were entered into the MetaboAnalyst Statistical Analysis module (v3.5), where PCA and PLS-DA analysis confirmed the presence of batch errors in each file²⁷. These data were then entered into the MetaboAnalyst Batch Effect Adjustment Utility (v3.5), and batch corrected via ComBat. Each file underwent BPCA imputation, and the lipid and aqueous files were combined into one master file. Duplicate metabolites found in both files were renamed to reflect their origin. The 9-hour test condition was removed, and t-tests were performed comparing metabolites from male

and female participants. Significant metabolites found from this test were excluded from analysis.

Data Analysis

Data from the baseline vs. 5-hour condition were entered into the MetaboAnalyst Biomarker Analysis Module (v4.0) and subjected to AUC analysis. The 25 metabolites with the greatest AUC scores were selected and entered into R (v3.4.3) for analysis. These data underwent binomial logistic regression using the “glm” function from the MASS package (v.7.3-48) in R^{61,62}. “Family” was set to “binomial” and “link” was set to “logit”, From this analysis, six significant metabolites were selected. A biomarker score was created for each sample using the following equation:

$$\beta^S + \sum_{k=1}^n \beta H$$

where β^S is the sex β -coefficient, β is each metabolite’s β -coefficient, and H is the peak height value for the corresponding metabolite, representing the relative abundance. The binomial logistic regression function provided these β -coefficients. This composite score was then subjected to AUC analysis in the MetaboAnalyst Biomarker module (v.4.0). A 50% classifier holdout test was also performed using the same set of data, where half of the class labels were randomly deleted using the Random Integer Set Generator on Random.org⁶³. These data were entered into the MetaboAnalyst Tester module (v4.0), and the biomarker score predicted the classification of samples with a missing label using both the PLS-DA and linear support vector machine algorithms. The biomarker was then tested in an independent dataset from a different protocol where participants were also subjected to insufficient sleep.

Independent Protocol Participants

We studied 36 participants (18 women) ages 25.5 ± 4.7 years. The Colorado Multiple Institutional Review Board and the Colorado Clinical Translational Science Institute's Scientific Advisory and Review Committee approved all protocols, and all participants gave informed written consent prior to the study. We determined the participants free of medical and psychological disorders using a variety of health screening procedures, including a physical examination, clinical psychiatric interview, electrocardiogram, metabolic panel, toxicology screen, and polysomnographic sleep disorders assessment. Inclusion criteria for the study included: 18-40 years old, a BMI between 18.5 and 24.9 kg/m², less than two days of exercise per week, a habitual sleep schedule between 7 and 9 hours, and living above 1,600 meters for three months prior to the study. Exclusion criteria included: medical and psychiatric conditions, clinically significant sleep disorders, drug use within one month of the study, uncorrected visual impairment, shift work during the prior year, blood donation in the prior 30 days, women with menstrual cycles lasting outside of the range from 25-32 days with more than 3 days variation in a given month, weight change of greater than 15 pounds 6 months prior to study, travel more than 1 time zone 3 weeks prior to the study, pregnancy, or nursing.

Independent Protocol Procedure

Participants stopped caffeine and alcohol consumption one week prior to the study, during which they maintained a consistent 9-hour sleep schedule based on their habitual schedule, monitored by wrist actigraphy, time-stamped call-ins, and sleep diaries. Three days prior to the study, participants ate an outpatient diet created by Clinical Translational Research Center nutritionists based on their resting metabolic rate data collected during the medical

screening. Participants stopped exercising during these three days. The first three days of the inpatient study consisted of baseline assessment, where participants maintained their habitual sleep schedule and outpatient diet while living in a laboratory environment exposed to natural sunlight, via a window, and 200 lux room lighting. After the baseline assessment, participants completed one of three randomly assigned treatment conditions: control, sleep restriction, or weekend recovery (Figure 3). The control condition consisted of 10 additional days of habitual sleep time; the sleep restriction condition comprised 10 additional days of 5 hour sleep opportunities followed by 3 days of recovery sleep; and the weekend recovery condition consisted of an additional 5 days of 5 hour sleep opportunities, then 2 nights of *ad libitum* sleep opportunities in bed (to mimic typical weekend sleep habits), then 3 days of 5 hour sleep opportunities, then recovery sleep. Participants were given *ad libitum* access to food during each experimental condition.

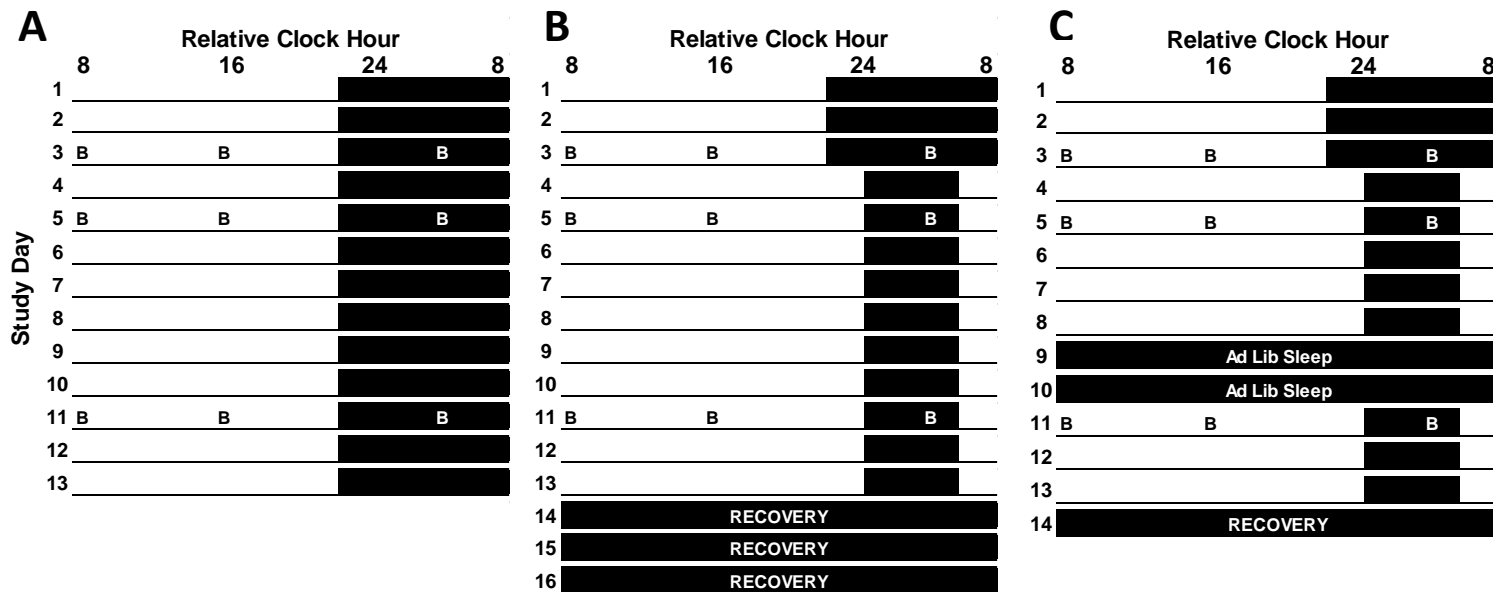


Figure 3. Independent Dataset Study Protocol. Participants were randomly assigned to one of three treatment groups: a control condition consisting of 10-hour sleep opportunities (A), a sleep restriction condition consisting of 5-hour sleep opportunities (B), and a weekend recovery condition consisting of 5 days of 5-hour sleep opportunities, followed by two nights of *ad libitum* sleep, followed by an additional two nights of 5-hour sleep opportunities (C). We collected blood during study days 3, 5, and 11 after 1 and 13 hours awake, and during the sleep mid-point via a catheter. Blood draws are marked with a “B,” and sleep opportunities are marked in black.

Independent Protocol Blood Collection and Processing

We collected blood during days 3, 5, and 11 during each protocol, at three time points: 1 hour awake, 13 hours awake, and at the sleep midpoint via in-arm catheter. The first sample each day was a fasted sample, drawn before the participant received breakfast. The sample preparations and mass spectrometry processing were identical to those described above in the “Blood Collection and Processing” section.

Independent Protocol Data Preparation

Lipid and aqueous mass spectrometer peak height analysis data files were filtered to remove metabolites with missing data found in 50% or more samples. Batch correction and sample imputation were identical to those described above in the “Data Preparation” section. The baseline day samples from the sleep restriction condition and the weekend recovery condition were excluded from analysis, and the post weekend recovery samples were excluded from the weekend recovery condition.

Independent Protocol Data Analysis

Equivalent metabolites for 4 of the 6 metabolites composing the biomarker were found in the independent protocol data based on generated annotations and mass and retention time differences. A biomarker score was created for each sample using the following equation:

$$\beta^S + \sum_{k=1}^n \beta H$$

where β is the β -coefficient for each metabolite from the “Data Analysis” section, and H is the corresponding peak height for that metabolite, representing the relative abundance. A receiver operating characteristic curve was created in the independent dataset between the control and

sleep restriction conditions, and an area under the curve score was created. Half of the class labels were then deleted at random based on numbers generated by the Random Integer Set Generator on Random.org. These data were entered into the MetaboAnalyst Tester module (v3.0), where the biomarker score was used to predict class assignments using both the PLS-DA and linear support vector machine algorithms. The accuracy of the prediction was calculated by comparing the predicted class assignments with the known class assignments.

Results

Biomarker Creation and Score

Six significant metabolites were obtained after performing binomial logistic regression on the 25 metabolites found using AUC in the baseline vs. 5-hour condition for the test dataset (Table 1). Sex was found to be a significant factor, with a β -coefficient of 2.52332 for males (0 for females) and a p value of 0.001053.

Table 1. Biomarker Metabolites with Logistic Regression with Probability Score and β coefficient

Metabolite	Probability	β Coefficient
(AQ) PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.000183***	-1.88333
Flavonol 3-O-alpha-L-rhamnosyl-1-6-beta-D-glucoside Esi+5.2899995	0.040019*	-0.95322
PC(18:4(6Z,9Z,12Z,15Z)/18:1(11Z))	0.015280*	-1.02159
Flavonol 3-O-alpha-L-rhamnosyl-1-6-beta-D-glucoside Esi+5.288002	0.034202*	0.80287
3,7,24-Trihydroxy-5-cholestanoyl-CoA	0.000833***	-0.24284
1239.4163@6.6779866	0.004924**	-0.15440

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Biomarker Score AUC

The receiver operating characteristic curve created using the biomarker score had an area under the curve of 0.943 with a 95% confidence interval of 0.903-0.974 which qualifies as an “excellent” score (Figure 4). This area under the curve was higher than all individual metabolite

scores (Table 2). The optimal biomarker score cut-off, maximizing sensitivity and specificity, was -52 (Figure 5).

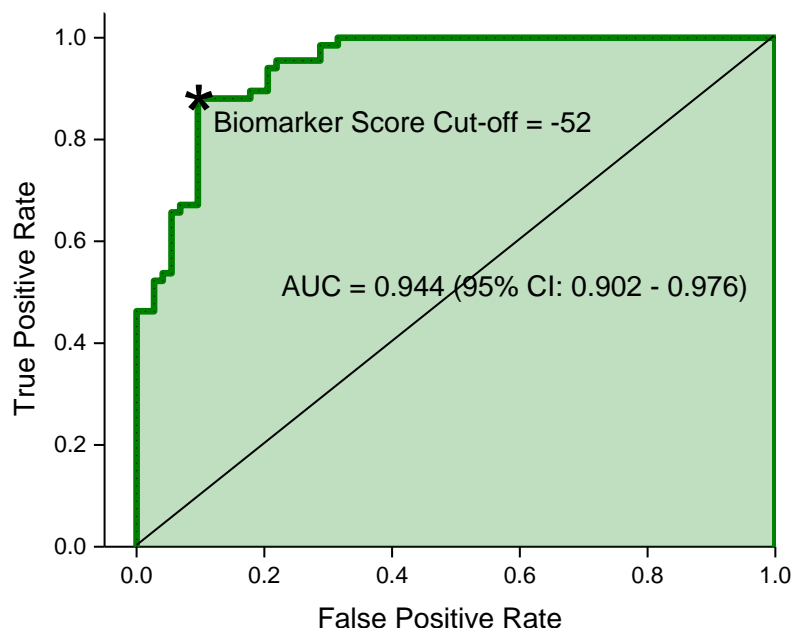


Figure 4. Biomarker Score Receiver Operator Characteristic Curve. Creating a univariate ROC using the biomarker score in the baseline and 5-hour condition yields an AUC of 0.944 with a 95% confidence interval of 0.902 – 0.976, indicating “excellent” performance. The ideal cut-off, optimizing for sensitivity and specificity, is -52, which results in a sensitivity of 0.881 and a specificity of 0.904.

Biomarker Score 50% Holdout Test

The biomarker score correctly predicted the class assignments of 91.4% of the holdout samples using the PLS-DA algorithm with 2 latent variables, and correctly predicted 86.4% of the holdout samples using the linear support vector machine algorithm.

Table 2. Receiver Operator Characteristic Curve AUC for Biomarker and Individual Component Metabolites

Metabolite	AUC	95% Confidence Interval
Biomarker Score	0.944	0.902-0.976
(AQ) PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.865	0.797-0.918
Flavonol 3-O-alpha-L-rhamnosyl-1-6-beta-D-glucoside Esi+5.2899995	0.835	0.756-0.902
PC(18:4(6Z,9Z,12Z,15Z)/18:1(11Z))	0.828	0.756-0.891
Flavonol 3-O-alpha-L-rhamnosyl-1-6-beta-D-glucoside Esi+5.288002	0.815	0.736-0.874
3,7,24-Trihydroxy-5-cholestanoyl-CoA	0.748	0.663-0.826
1239.4163@6.6779866	0.753	0.660-0.832

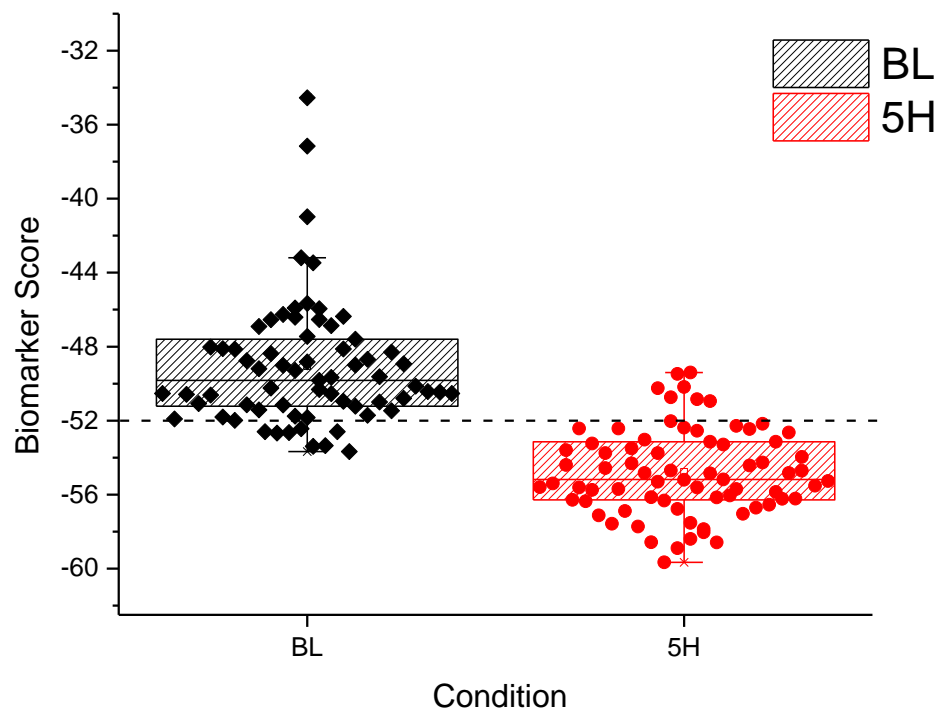


Figure 5. Biomarker Scores for the Baseline and 5-hour Conditions. Each created biomarker score clusters with other members of their condition. The average baseline biomarker score is -49.05, and the average 5-hour biomarker score is -54.78. The optimal cut-off score, maximizing specificity and sensitivity, is -52, marked with the dashed line.

Independent Protocol Biomarker Creation and Score

Where possible, equivalent biomarker metabolites in the independent dataset to the ones composing were used to create the verification biomarker (Table 3). The metabolite PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z)) had an identical annotation in the independent dataset. Flavonol 3-O-alpha-L-rhamnosyl-1-6-beta-D-glucoside had an identical annotation but a slightly different retention time value. No identical annotation was found for the two verification metabolites PE(19:1(9Z)/20:4(5Z,8Z,11Z,14Z)) and 1239.4108@7.008028, but they exhibited low mass and retention time differences with their metabolite counterparts in the test biomarker (Table 3). No equivalent metabolite was found for 3,7,24-Trihydroxy-5-cholestanoyl-CoA.

Table 3. Equivalent Biomarker Metabolites in Independent Dataset with Mass and Retention Time Differences with Original Biomarker Metabolites

Biomarker Metabolite	Independent Dataset Equivalent	Δ Mass	Δ Retention Time
(AQ) PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	(AQ) PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.0013	-0.0139868
Flavonol 3-O-alpha-L-rhamnosyl-1-6-beta-D-glucoside Esi+5.2899995	Flavonol 3-O-alpha-L-rhamnosyl-1-6-beta-D-glucoside	-0.001	0.184
PC(18:4(6Z,9Z,12Z,15Z)/18:1(11Z))	PE(19:1(9Z)/20:4(5Z,8Z,11Z,14Z))*	-0.0001	-0.0150114
Flavonol 3-O-alpha-L-rhamnosyl-1-6-beta-D-glucoside Esi+5.288002	Flavonol 3-O-alpha-L-rhamnosyl-1-6-beta-D-glucoside	-0.001	0.184
3,7,24-Trihydroxy-5-cholestanoyl-CoA	-	-	-
1239.4163@6.6779866	1239.4108@7.008028*	-0.0055	0.3300414

*Not exact match

Independent Protocol Biomarker Score AUC

The biomarker score was tested in the control vs. weekend recovery conditions of the independent dataset. The area under the receiver operating characteristic curve using the biomarker verification score was 0.781 with a 95% confidence interval of 0.673-0.870, qualifying as a “fair” score (Figure 6). This area under the curve was higher than all component metabolite’s area under the curve values (Table 4). The optimal cut-off, maximizing both sensitivity and specificity, was -0.599 (Figure 7).

Table 4. Receiver Operator Characteristic Curve AUC Values and Confidence Intervals for Verification Biomarker and Individual Component Metabolites

Metabolite	AUC	95% Confidence Interval
Biomarker Score	0.781	0.686-0.865
Flavonol 3-O-alpha-L-rhamnosyl-1-6-beta-D-glucoside	0.694	0.581-0.784
PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.617	0.510-0.713
1239.4108@7.008028	0.589	0.491-0.682
(AQ) PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.505	0.456-0.629

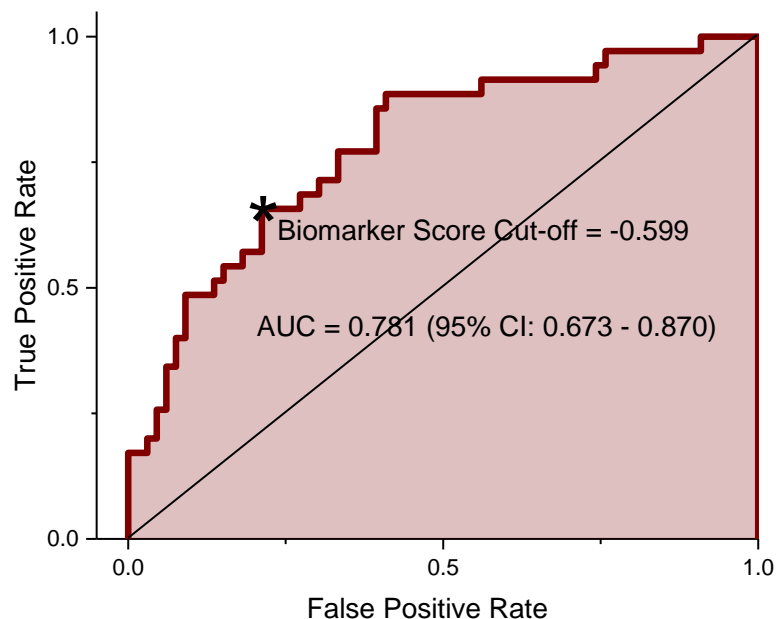


Figure 6. Verification Biomarker Receiver Operating Characteristic Curve. Creating a univariate ROC curve using the verification biomarker score in the control and weekend recovery sleep conditions in the independent dataset yields an area under the curve of 0.781 with a 95% confidence interval of 0.636-0.802. The optimal cut-off, maximizing both sensitivity and specificity, is -0.599, resulting in a sensitivity of 0.8 and a specificity of 0.7.

Independent Protocol Biomarker Score 50% Holdout Test

The verification biomarker score correctly predicted the class assignments of 64.7% of the holdout samples with the PLS-DA algorithm with 2 latent variables and the Linear Support Vector Machine algorithm.

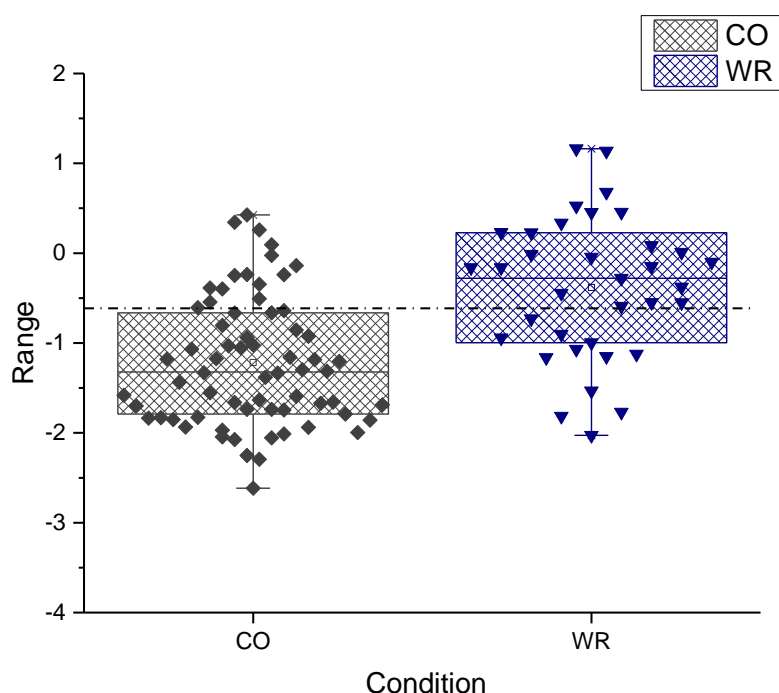


Figure 7. Biomarker Scores for the Control and Weekend Recovery Conditions in the Independent Dataset. The average control biomarker score is -1.22, and the average sleep restriction biomarker score is -0.38. The optimal cut-off score, maximizing specificity and sensitivity, is -0.599, marked with the dashed line.

Discussion

The creation of a serum biomarker which holds a high predictive accuracy in classifying unknown samples based on sleep condition supports the feasibility of the future use of such a tool to diagnose sleep status. Using several significant metabolites to build a biomarker score provides greater discrimination between sufficient and insufficient sleep conditions than individual metabolites can, suggesting this method is optimal for developing a robust biomarker. Creating a biomarker score using β -coefficients weights each metabolite based on its contribution to the score. This suggests metabolites with a greater magnitude β -coefficient, such as PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z)) with its β -coefficient of -1.88333, potentially has more physiological relevance to sleep than 1239.4163@6.6779866 with its β -coefficient of -0.15440.

A dearth of identified metabolites remains a problem in metabolomics; the vast quantity of metabolites present in the human body precludes easy identification and classification, and increased mass spectrometer power continues to uncover novel metabolites. As such, a minority of metabolites hold entries in databases such as the Kyoto Encyclopedia of Genes and Genomes or Human Metabolome Database, meaning most metabolites detected by the mass spectrometer remain unannotated, such as 1239.4163@6.6779866, or partially annotated, such as PC(18:4(6Z,9Z,12Z,15Z)/18:1(11Z)). Identifying these uncharacterized metabolites is critical to develop a greater understanding of the biochemical effects of insufficient sleep. To accomplish this identification, researchers can perform targeted analysis of the metabolites using techniques such as tandem mass spectrometry to elucidate its structure⁶⁴. Characterizing unknown metabolites is an essential step before further clinical development, as it can reveal insights into

the disease mechanism, and allow investigators to develop targeted assays for metabolites of interest.

Examining the metabolites contained in the biomarker opens some interesting hypothesis for future testing. According to the Kyoto Encyclopedia of Genes and Genomes, 3,7,24-Trihydroxy-5-cholestanoyl-CoA is involved in the primary bile acid biosynthesis pathway⁶⁰. Ferrell and Chiang describe how short term circadian disruption impairs bile synthesis pathways in mice, especially when their subjects consumed with a western diet equivalent composed of 42% fat, 42.7% carbohydrate, and 15.2% protein⁶⁵. Their western diet contains 12% more fat, 12.3% less carbohydrates, and 0.2% less protein than the diet received by participants in our study. However, if human biosynthesis pathways respond similarly to those found in mice, and the exact macronutrient composition of the diet is not a confounding factor, perhaps bile synthesis metabolites are a strong candidate for insufficient sleep biomarker development. Metabolite databases contain little information on the remainder of the metabolites, preventing further functional investigation.

The biomarker scores created in this thesis were all optimized for high sensitivity and specificity. However, future clinical applications may benefit from optimizing one over the other. Specificity refers to this biomarker's ability to correctly discriminate people who are sleeping enough, and sensitivity refers to this biomarker's ability to correctly discriminate people who are not sleeping enough⁶⁶. In the case of insufficient sleep, a clinician may want a biomarker with higher specificity than sensitivity. Incorrectly diagnosing an individual as not sleeping enough has less of a negative impact than incorrectly diagnosing an individual as sleeping enough; the individual in the first case might be given strategies and tools to try and sleep more, which has fewer negative health risks than not catching someone who experiences chronic levels

of insufficient sleep⁶⁷. Such exact tuning of a biomarker depends on further verification of the biomarker's performance.

Much work remains before a biomarker for insufficient sleep will make its way into the clinic. The study protocol used in this process was not designed for biomarker analysis, and as such introduce confounds which could potentially diminish the strength of the biomarker. For example, during the 5-hour experimental condition, males ate 68% more calories and women ate 19% more calories than their weight maintenance level; these differences in consumption may have influenced the metabolome, and thus altered the biomarker⁶⁸. Gordon et al. describe a method to address similar confounds which change based on sex, which I implemented, but the influence of the confound cannot be ignored⁶⁹. A biomarker which discriminates sleep status based on overeating may function correctly, but it is not sensitive to the underlying physiological changes of insufficient sleep, and thus may not perform optimally in catching those who experience insufficient sleep but do not overeat. To investigate overeating's effect on the biomarker, I separated the fasting and fed samples for the original and independent datasets and performed receiver operator characteristic curve and sample holdout tests on each. A two-tailed paired t-test performed on various fasted and fed AUC values for both datasets suggests the differences between the groups are not statistically significant, with a p value of 0.2933. These data indicate this biomarker holds some resistance to fasting or fed status, but future experiments should control food intake.

External biomarker verification is an important step in clinical development. This thesis tested the biomarker in an independent dataset using a population also experiencing insufficient sleep conditions. Those participants were randomly assigned into one of three conditions: a control condition, a weekend recovery condition, or a sleep restriction condition, and blood was

taken during three days of the experiment. The control and weekend recovery conditions were used as the two sample groups to test the biomarker in, and the baseline blood sample data was removed from the weekend recovery condition, as the participants received sufficient sleep during that time point. Structural data issues in the sleep restriction condition necessitated using the weekend recovery condition for analysis. While neither the receiver operator characteristic curve value of 0.781, classified as “fair,” nor the holdout verification score of 64.7% correct, are as high as their corresponding scores in the original dataset, some concessions were made while translating the score to the new dataset. The mass spectrometer did not detect all metabolites used in the original biomarker in the independent dataset, meaning the converted biomarker contains less data, potentially giving it less predictive value. Additionally, two of the metabolites chosen in the verification biomarker, PE(19:1(9Z)/20:4(5Z,8Z,11Z,14Z)) and 1239.4108@7.008028, closely matched the mass and retention time of their corresponding metabolites, PC(18:4(6Z,9Z,12Z,15Z)/18:1(11Z)) and 1239.4163@6.6779866 respectively, but not their annotation. A reasonable chance exists that these chosen metabolites are not analogous to their original metabolites, and thus hold weaker predictive value, further harming the biomarker’s performance. Taking these limitations into account, the AUC and verification percentage are stronger than they first appear. Future investigations should ensure the same metabolites are found in both the test and independent dataset, to ensure an accurate assessment of the model’s performance.

Future clinical development would also require verification in an external population living outside laboratory conditions. The controlled laboratory environment is important to reduce biological noise and develop a biomarker sensitive to a specific condition. However, a biomarker which only works in laboratory conditions holds no value for clinical use, where

individuals experience a range of environments, sleep schedules, and dietary habits prior to entering the clinic, all of which influence the metabolome⁷⁰⁻⁷³. Thus, testing the biomarker in a free-living human population is a vital step in bringing a potential biomarker to the clinic. Assay development forms another step towards preparing a biomarker for a clinical use. The present biomarker uses arbitrary peak height and biomarker score values to discriminate conditions. In the future, we would need to develop a targeted assay specific for the biomarker metabolites. This process could allow us to establish metabolite concentration thresholds in the blood corresponding to different levels of insufficient sleep. To this end, the relative upregulation or downregulation of the metabolites of interest should be investigated, to allow the creation of an effective assay.

This thesis describes a method to create a biomarker sensitive to insufficient sleep using metabolomic techniques. An ongoing study in our laboratory is designed for biomarker creation and analysis, granting both another dataset to verify this biomarker in and opportunity to develop another biomarker building off the methods described here. That study controls for food intake over the course of the entire study, ideally preventing any potential confounds with a participant's fed or fasted state. Future advances in mass spectrometry technologies will further expand the range of detectable metabolites, increasing the probability of detecting metabolites of interest for future studies.

Acknowledgments

Special thanks to Dr. Kenneth Wright Jr., my thesis advisor, for providing an opportunity to conduct research in his laboratory. I could not have completed this project without the help of Dr. Christopher Depner, who acted as my mentor during this project and provided extensive guidance and assistance towards the development of the project. Thank you to Dr. Pamala Harvey and Dr. Bradley Olwin for taking the time to serve on my thesis committee. I am grateful for the past and present members of the Sleep and Chronobiology Laboratory for supporting me and making this research possible, and for the Undergraduate Research Opportunities Program for funding me. Finally, I would like to thank my friends, family, and my castmates at the Department of Theatre and Dance for keeping me sane over the past several years.

References

1. World Health Organization & International Programme on Chemical Safety. Biomarkers in risk assessment: validity and validation. *Environ. Heal.* 113 (2001).
2. Voultsios, A., Kennaway, D. J. & Dawson, D. Salivary melatonin as a circadian phase marker: validation and comparison to plasma melatonin. *J. Biol. Rhythms* **12**, 457–466 (1997).
3. Liu, Y. *et al.* Prevalence of healthy sleep duration among adults — United States, 2014. *MMWR. Morb. Mortal. Wkly. Rep.* **65**, 137–141 (2016).
4. Nathaniel, W., Badr, M. S., Belenky, G., Bliwise, D. L. & Buxton, O. M. SLEEP - recommended amount of sleep for a healthy adult: A joint consensus statement of the american academy of sleep medicine and sleep research society. *Sleep* **38**, 843–844 (2015).
5. Shan, Z. *et al.* Sleep duration and risk of type 2 diabetes: A meta-analysis of prospective studies. *Diabetes Care* **38**, 529–537 (2015).
6. Kakizaki, M. *et al.* Sleep duration and the risk of breast cancer: The Ohsaki Cohort study. *Br. J. Cancer* **99**, 1502–1505 (2008).
7. Neckelmann, D., Mykletun, A. & Dahl, A. a. Chronic insomnia as a risk factor for developing anxiety and depression. *Sleep* **30**, 873–880 (2007).
8. Curcio, G., Ferrara, M. & De Gennaro, L. Sleep loss, learning capacity and academic performance. *Sleep Medicine Reviews* **10**, 323–337 (2006).
9. Miyata, S. *et al.* Insufficient sleep impairs driving performance and cognitive function. *Neurosci. Lett.* **469**, 229–233 (2010).
10. Colten, H., Altevogt, B. B. M. & Colten., H. *Sleep disorders and sleep deprivation: an*

unmet public health problem. Committee on Sleep Medicine and Research (2006).

doi:10.1097/01.CHI.0000270812.55636.3b

11. Kapur, V. *et al.* The medical cost of undiagnosed sleep apnea. *Sleep* **22**, 749–755 (1999).
12. Filip, I. *et al.* Public health burden of sleep disorders: underreported problem. *Journal of Public Health (Germany)* **25**, 243–248 (2017).
13. Van Den Berg, J. F. *et al.* Disagreement between subjective and actigraphic measures of sleep duration in a population-based study of elderly persons. *J. Sleep Res.* **17**, 295–302 (2008).
14. Short, M. A., Gradisar, M., Lack, L. C., Wright, H. & Carskadon, M. A. The discrepancy between actigraphic and sleep diary measures of sleep in adolescents. *Sleep Med.* **13**, 378–384 (2012).
15. Xia, J., Broadhurst, D. I., Wilson, M. & Wishart, D. S. Translational biomarker discovery in clinical metabolomics: An introductory tutorial. *Metabolomics* **9**, 280–299 (2013).
16. Kasukawa, T. *et al.* Human blood metabolite timetable indicates internal body time. *Proc. Natl. Acad. Sci.* **109**, 15036–15041 (2012).
17. Weljie, A. M. *et al.* Oxalic acid and diacylglycerol 36:3 are cross-species markers of sleep debt. *Proc. Natl. Acad. Sci.* (2015). doi:10.1073/pnas.1417432112
18. Psychogios, N. *et al.* The human serum metabolome. *PLoS One* **6**, e16957 (2011).
19. Knutson, K. L. Impact of sleep and sleep loss on glucose homeostasis and appetite regulation. *Sleep Medicine Clinics* **2**, 187–197 (2007).
20. Van Cauter, E. *et al.* Modulation of glucose regulation and insulin secretion by circadian rhythmicity and sleep. *J. Clin. Invest.* **88**, 934–942 (1991).
21. Griffiths, W. J. *et al.* Targeted metabolomics for biomarker discovery. *Angewandte*

- Chemie - International Edition* **49**, 5426–5445 (2010).
22. Vinayavekhin, N. & Saghatelian, A. Untargeted metabolomics. *Current Protocols in Molecular Biology* (2010). doi:10.1002/0471142727.mb3001s90
 23. Wang, P., Whiteaker, J. R. & Paulovich, A. G. The evolving role of mass spectrometry in cancer biomarker discovery. *Cancer Biol. Ther.* **8**, 1083–1094 (2009).
 24. March, R. E. An introduction to quadrupole ion trap mass spectrometry. *Journal of Mass Spectrometry* **32**, 351–369 (1997).
 25. Wishart, D. S. *et al.* HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
 26. Smith, C. A. *et al.* METLIN: A metabolite mass spectral database. in *Therapeutic Drug Monitoring* **27**, 747–751 (2005).
 27. Xia, J., Sinelnikov, I. V., Han, B. & Wishart, D. S. MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Res.* (2015). doi:10.1093/nar/gkv380
 28. Holmes, E. & Antti, H. Chemometric contributions to the evolution of metabonomics: mathematical solutions to characterising and interpreting complex biological NMR spectra. *Analyst* **127**, 1549–1557 (2002).
 29. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: A basic tool of chemometrics. in *Chemometrics and Intelligent Laboratory Systems* **58**, 109–130 (2001).
 30. Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **2**, 37–52 (1987).
 31. Bylesjo, M. *et al.* OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* **20**, 341–351 (2006).
 32. Wehrens, R. *et al.* Improved batch correction in untargeted MS-based metabolomics.

- Metabolomics* **12**, (2016).
33. Fare, T. L. *et al.* Effects of atmospheric ozone on microarray data quality. *Anal. Chem.* **75**, 4672–4675 (2003).
 34. Casella, G. An introduction to empirical bayes data analysis. *Am. Stat.* **39**, 83–87 (1985).
 35. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
 36. Kirwan, J. A., Broadhurst, D. I., Davidson, R. L. & Viant, M. R. Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. *Anal. Bioanal. Chem.* **405**, 5147–5157 (2013).
 37. Yu, T., Ye, H., Chen, Z., Ziober, B. L. & Zhou, X. Dimension reduction and mixed-effects model for microarray meta-analysis of cancer. *Front Biosci* **13**, 2714–2720 (2008).
 38. Payne, T. G., Southam, A. D., Arvanitis, T. N. & Viant, M. R. A signal filtering method for improved quantification and noise discrimination in fourier transform ion cyclotron resonance mass spectrometry-based metabolomics data. *J. Am. Soc. Mass Spectrom.* **20**, 1087–1095 (2009).
 39. Hrydziuszko, O. & Viant, M. R. Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. *Metabolomics* **8**, 161–174 (2012).
 40. Gromski, P. *et al.* Influence of missing values substitutes on multivariate analysis of metabolomics data. *Metabolites* **4**, 433–452 (2014).
 41. Xia, J. & Wishart, D. S. Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis. in *Current Protocols in Bioinformatics* (2016). doi:10.1002/cpbi.11
 42. Kenny, L. C. *et al.* Robust early pregnancy prediction of later preeclampsia using

- metabolomic biomarkers. *Hypertension* **56**, 741–749 (2010).
43. Wei, R. *et al.* Missing value imputation approach for mass spectrometry-based metabolomics data. *Sci. Rep.* **8**, (2018).
 44. de Brevern, A. G., Hazout, S. & Malpertuy, A. Influence of microarray experiments missing values on the stability of gene groups by hierarchical clustering. *BMC Bioinformatics* **5**, (2004).
 45. Nounou, M. N., Bakshi, B. R., Goel, P. K. & Shen, X. Bayesian principal component analysis. *J. Chemom.* **16**, 576–595 (2002).
 46. Brock, G. N., Shaffer, J. R., Blakesley, R. E., Lotz, M. J. & Tseng, G. C. Which missing value imputation method to use in expression profiles: A comparative study and two selection schemes. *BMC Bioinformatics* **9**, (2008).
 47. Oba, S. *et al.* A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics* **19**, 2088–2096 (2003).
 48. Bamber, D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.* **12**, 387–415 (1975).
 49. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
 50. Obuchowski, N. A., Lieber, M. L. & Wians, F. H. ROC curves in clinical chemistry: uses, misuses, and possible solutions. *Clinical Chemistry* **50**, 1118–1125 (2004).
 51. Zweig, M. H. & Campbell, G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry* **39**, 561–577 (1993).
 52. Søreide, K. Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. *J. Clin. Pathol.* **62**, 1–5 (2009).

53. Kleinbaum, D. G. & Klein, M. *Logistic Regression. Statistics for Biology and Health* (2010). doi:10.1007/978-1-4419-1742-3
54. Edelstein, C. L. *Biomarkers of kidney disease. Biomarkers of Kidney Disease* (2011). doi:10.1016/C2009-0-60986-7
55. Hori, S. S. & Gambhir, S. S. Mathematical model identifies blood biomarker-based early cancer detection strategies and limitations. *Sci. Transl. Med.* **3**, (2011).
56. Leichtle, A. B. *et al.* Pancreatic carcinoma, pancreatitis, and healthy controls: Metabolite models in a three-class diagnostic dilemma. *Metabolomics* **9**, 677–687 (2013).
57. Rice, L. M. *et al.* A longitudinal biomarker for the extent of skin disease in patients with diffuse cutaneous systemic Sclerosis. *Arthritis Rheumatol.* **67**, 3004–3015 (2015).
58. Mamas, M., Dunn, W. B., Neyses, L. & Goodacre, R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Arch. Toxicol.* **85**, 5–17 (2011).
59. Fahy, E. *et al.* Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* **50**, S9–S14 (2009).
60. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
61. R Development Core Team. R: a language and environment for statistical computing. *R Found. Stat. Comput. Vienna Austria* **0**, {ISBN} 3-900051-07-0 (2016).
62. Kafadar, K., Koehler, J. R., Venables, W. N. & Ripley, B. D. Modern Applied Statistics with S-Plus. *Am. Stat.* **53**, 86 (1999).
63. Haahr, M. & Haahr, S. Random.org. *Randomness and Integrity Services Ltd.* (2015).

Available at: <https://www.random.org/>.

64. Bowen, B. P. & Northen, T. R. Dealing with the unknown: metabolomics and metabolite atlases. *J. Am. Soc. Mass Spectrom.* **21**, 1471–1476 (2010).
65. Ferrell, J. M. & Chiang, J. Y. L. Short-Term circadian disruption impairs bile acid and lipid homeostasis in mice. *C. Cell. Mol. Gastroenterol. Hepatol.* **1**, 664–677 (2015).
66. Lalkhen, A. G. & McCluskey, A. Clinical tests: sensitivity and specificity. *Contin. Educ. Anaesthesia, Crit. Care Pain* **8**, 221–223 (2008).
67. Léger, D., Beck, F., Richard, J. B., Sauvet, F. & Faraut, B. The risks of sleeping ‘too much’. Survey of a national representative sample of 24671 adults (INPES health barometer). *PLoS ONE* **9**, (2014).
68. Markwald, R. R. *et al.* Impact of insufficient sleep on total daily energy expenditure, food intake, and weight gain. *Proc. Natl. Acad. Sci.* (2013). doi:10.1073/pnas.1216951110
69. Gorden, D. L. *et al.* Biomarkers of NAFLD progression: a lipidomics approach to an epidemic. *J. Lipid Res.* **56**, 722–36 (2015).
70. Chen, C. H. S., Yuan, T. H., Shie, R. H., Wu, K. Y. & Chan, C. C. Linking sources to early effects by profiling urine metabolome of residents living near oil refineries and coal-fired power plants. *Environ. Int.* **102**, 87–96 (2017).
71. Wang, X. *et al.* Serum metabolome biomarkers associate low-level environmental perfluorinated compound exposure with oxidative /nitrosative stress in humans. *Environ. Pollut.* **229**, 168–176 (2017).
72. Winnike, J. H., Busby, M. G., Watkins, P. B. & O’Connell, T. M. Effects of a prolonged standardized diet on normalizing the human metabolome. *Am. J. Clin. Nutr.* **90**, 1496–1501 (2009).

73. Medina, S., Domínguez-Perles, R., Ferreres, F., Tomás-Barberán, F. A. & Gil-Izquierdo, Á. The effects of the intake of plant foods on the human metabolome. *TrAC - Trends in Analytical Chemistry* **52**, 88–99 (2013).