Pilot Study: Metabolic and microbial profiling of lung cancer
Wake Forest Baptist Comprehensive Cancer Center (WFBCCC)
WFBCCC # 03219


**Principal Investigator:** Andrew Bishop, PhD, Assistant Professor
Internal Medicine/ Molecular Medicine
Wake Forest Baptist Health
Medical Center Blvd
Winston Salem, NC 27157

**Molecular Medicine Co-Investigator(s):** Kimberly Reeves, PhD, Assistant Professor
Internal Medicine/ Molecular Medicine

Tony Reeves, PhD, Assistant Professor
Internal Medicine/ Molecular Medicine

**Pulmonary Co-Investigator(s):** Travis Dotson, MD, Assistant Professor
Internal Medicine/ Pulmonary and Critical Care

Christina Bellinger, MD, Assistant Professor
Internal Medicine/ Pulmonary and Critical Care

**Cardiothoracic surgery Co-Investigator(s):** Leonard (Jim) Wudel, MD, Associate Professor
Surgical Sciences/ Cardiothoracic surgery

**Biostatistician:** Ralph D'Agostino Jr. PhD, Professor
Biostatistical Sciences/ Comprehensive Cancer Center

**Study Team:** Ashley Wudel, RN, Clinical Coordinator

Sharon Averill, RN, Clinical Research Nurse

Justin Prescott, Graduate student

**Regulatory Contact:** Emily Teal, MA
Comprehensive Cancer Center

**Protocol Editor:** Megan Farmer, MS
Comprehensive Cancer Center

Pilot Study: Metabolic and microbial profiling of lung cancer
Wake Forest Baptist Comprehensive Cancer Center (WFBCCC)
WFBCCC # 03219

**Participating Institution(s):** Wake Forest Baptist Health

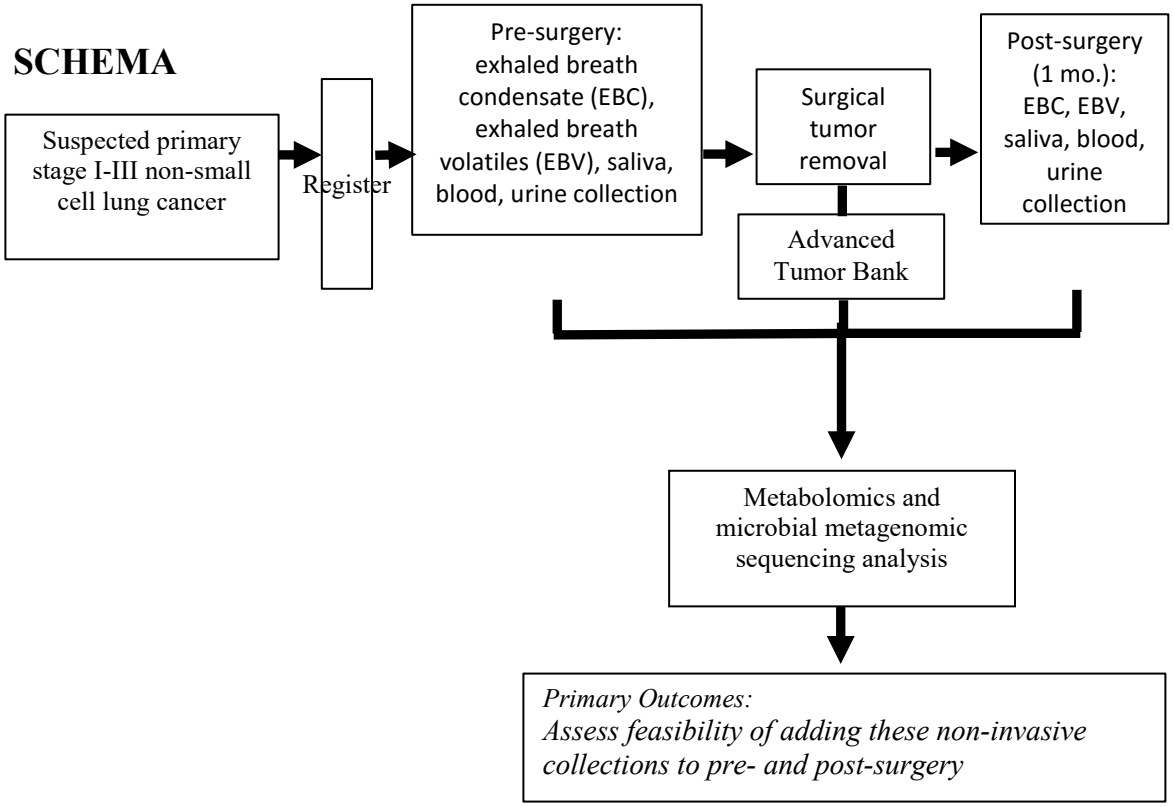
**Version Date:** 09/03/2019
**Amended:**

**ClinicalTrials.gov:** **NCT03998189**

**Confidential**

# Table of Contents

**SCHEMA**




*Secondary Outcomes:*
*Assess changes in metabolites and microbial populations between lung cancer patients' biosamples pre- and post-surgery*

*Identification of lung cancer stage specific signatures*
*Identify signatures which associate with pulmonary function*

# 1.0 Introduction and Background

Lung cancer (LC) is the second most common type of cancer and leading cause of cancer-related deaths in the United States (1). It is estimated that there will be over 120,000 new cases in 2018 (1). Early detection leads to earlier treatment, and early LC stage detection has clear benefit in survival rates(2). Clinical tools for diagnostic screening of LC are limited, but efforts have been made to develop molecular techniques to identify LC during bronchoscopy procedures (3). The National Lung Screening Trial screened patients at high-risk of LC by computed tomography (CT) scans (4). Mortality rates were reduced through the screening, but a number of challenges to CT screening were noted. For example, one in four patients showed incidental abnormalities in the lung potentially not representing LC, and further invasive tests including biopsies and/or surgery were required to rule out cancer(4). Additionally, the uncertainty surrounding incidental findings is suspected to lead to further patient stress, which may further impact the patient's health.

It is apparent that complex diseases, like cancer, require powerful and sensitive technologies to uncover underlying mechanisms of the disease and even to detect the presence of the disease. A metabolomics approach is able to obtain a snapshot of biological and molecular mechanisms underlying a disease process by capturing metabolic profiles that are indicative of cellular metabolism. Current profiling technology results in the identification of hundreds to thousands of metabolites, which can be used as an unbiased analysis of biochemical pathways and potentially target biological dysfunction(5) (6) (7). A number of studies have utilized metabolomics approaches to profile LC (8). Many of these studies focused on tumor versus non-tumor tissue or serum in cancer patients versus healthy controls. We seek to assess metabolic signatures in samples from patients before and after surgical removal of their tumor.

Recent studies have also focused on the link between cancer and the human microbiota, with special interest in determining how the oral microbiota contributes to the pathogenesis of various cancer types, including LC(9). It is well-established that the lungs are not sterile, and scientists have recently demonstrated that the lung microbiome is very similar to the oral microbiome(10); (11); (12). Furthermore, periodontal disease is associated with LC risk(13); (14), suggesting an association between the oral microbiome and LC risk, but further studies are needed to determine how oral microbial variations contribute to LC(9). The most informative approach to assessing the human microbiota is through shotgun metagenomic sequencing, which allows the characterization of microorganisms down to the species or even strain level and provides information on the putative function of the microbiota. Collecting saliva before and after surgical resection offers a noninvasive approach for studying the role oral microbiota play in LC, which has not previously been done.

Utilizing these unbiased approaches will be pivotal to understanding the molecular dysregulation in LC and will potentially lead to the discovery of novel signatures within non-invasive biosamples that can be used as a diagnostic tool for identifying and characterizing LC. Our long-term goal is to translate findings from this study into diagnostics for the clinic, giving clinicians robust information to identify at-risk patients prior to the observation of symptoms. The need for a non-invasive, low-risk, highly-specific diagnostic tool for LC is still evident. By profiling patients prior to and post-surgical resection, we will identify potential molecular markers that may be indicative of LC.

# 2.0 Objectives

This pilot study will establish non-invasive sample collections, including breath, saliva, blood and urine pre-surgery and at the patient's one-month post-surgery follow-up visit. Patients with suspected non-small cell lung cancer (NSCLC) stage I-III will be recruited from Co-I Jim Wudel's clinic.

### 2.1 Primary Objective

2.1.1 To evaluate the feasibility of adding non-invasive sample collections in the pre-surgical setting and at the post-surgery follow-up visit.

### 2.2 Secondary Objective

2.2.1 To identify and assess metabolic and microbial signatures collected at pre- and post-surgery and determine which are indicative of LC.

2.2.2 To identify signatures which are associated with LC stage

2.2.3 To identify signatures which are impacted by patient's pulmonary function status

## 3.0 Study Population

### 3.1 Inclusion Criteria

3.1.1 Male and female patients age ≥18 years, of all racial and ethnic origins, with suspected NSCLC stages I, II, and III, as evident through radiographic evidence and felt acceptable to undergo surgical resection.

3.1.2 Patients who have the ability to understand and the willingness to sign a written consent form.

### 3.2 Exclusion Criteria

3.2.1 Patients who are have taken antibiotics within two weeks.

3.2.2 Patients who are on continuous supplemental oxygen.

3.2.3 Patients currently undergoing active treatment for other malignancies.

3.2.4 Subjects who are unable or unwilling to provide consent.

### 3.3 Inclusion of Women and minorities

3.3.1 Women and men of all races and ethnicity who meet the above-described eligibility criteria are eligible for this trial.

~~3.3.2~~ The study consent form will also be provided in Spanish for Spanish-speaking participants. Based on WFBCCC population estimates, approximately 51% of participants to will be women. Translating this to our sample size estimate of 90, we plan to enroll 45 women and 45 men. Similarly, we expect approximately 6.6% of study participants to be Hispanic/Latino (N=6), 13.3% Black or African

American (N=12), 0.48% American Indian/Alaska Native (N=1), and 1.8% Asian (N=1).

# 4.0 Methods

## 4.1 Registration Procedures

All patients entered into any WFBCCC trial, whether treatment, companion, or cancer control trial, **must** be linked with a study protocol in EPIC within 24 hours of Informed Consent. Patients **must** be registered prior to the initiation of the study.

You must perform the following steps in order to ensure prompt registration of your patient:

1.0 Complete the Eligibility Checklist (Appendix A).
2.0 Complete the Protocol Registration Form (Appendix B).
3.0 Alert the Cancer Center registrar by phone, *and then* send the signed Informed Consent Form, Eligibility Checklist and Protocol Registration Form to the registrar, either by fax or e-mail.

*Protocol Registration is open from 8:30 AM - 4:00 PM, Monday-Friday.

4.0 Fax/e-mail ALL eligibility source documents with registration. Patients **will not** be registered without all required supporting documents.

Note: If labs were performed at an outside institution, provide a printout of the results. Ensure that the most recent lab values are sent.

To complete the registration process, the Registrar will:

- assign a patient study number
- register the patient on the study

## 4.2 Experimental or Data Collection Methods

### 4.2.1 Clinical data collection:

All patients will be recruited within the Wake Forest Medical Center.
Medical history will be obtained from the patients charts to identify demographics (age, race, ethnicity, zip code), oncologic history (date of diagnosis and cancer type), and medical history of disease. Pulmonary function tests (PFTs) will be performed by Wake Forest Baptist Health Pulmonary Function Lab as a standard of care for pre-surgical assessment. PFTs will be accepted from other certified laboratories within six months of their appointmen. Pulmonary function is

defined by undergoing standard pulmonary function studies per American Thoracic Society recommendations. The collection of these pulmonary function studies is often done as part of the pre-surgical assessment collected by a cardiothoracic surgical team. Also, the collection of pulmonary functions studies can help with the classification and treatment of underlying obstructive lung disease that can be crucial to post-operative management.

At pre- and post- surgery, the research nurse will obtain current medications, food/ingestion survey and tobacco use information. This information will be necessary to identify any exogenous factors confounding both metabolic and microbial signatures. Following surgery, information from the patient's final pathology report will be obtained to provide information on the tumor site location, characteristics of the cancer removed, and stage of the cancer.

Figure 1: ReCIVA breath collection device

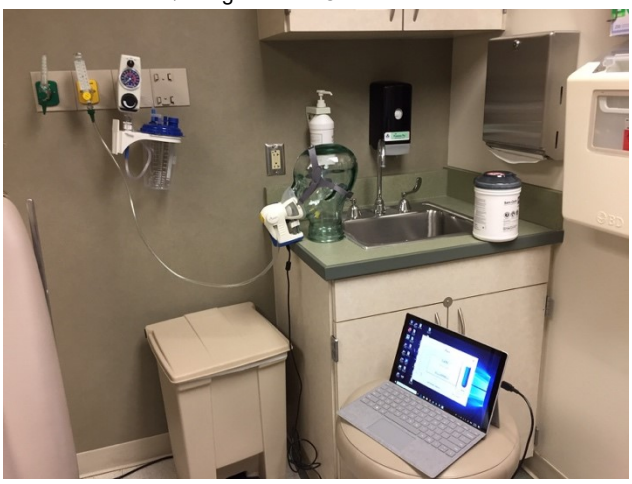


Figure 2: Clinic breath gas collection setup

**4.2.2 Non-invasive sample collections:**

**Pre-surgery follow-up collection:** During the patient's surgery consultation meeting with Dr. Wudel, Upon entering the are the patient will be given a receptacle for collection of a urine specimen. Given that blood collection may be seen as invasive to the patient, the patient will be prompted for a blood sample. If the patient agrees, the research nurse will collect 1 redtop vacutainer serum tube and 1 purple top blood tube from the patient. We will then initiate the saliva and exhaled breath volatiles (EBV) and exhaled breath condensate (EBC) collection. For the saliva collection, patients will be given a saliva collection receptacle. This receptacle (SalivaBio – Passive Drool) allow for sterile transfer of saliva from the patient's mouth into collection tubes. For exhaled breath collection, we will utilize the ReCIVA (Figure 1, Owlstone medical), breath collection device. The device will be attached to the in-house medical air supply (yellow) with airflow set to 40L/min per the company's recommendation (Figure 2). One-time use masks will be placed over the patient's mouth and nose. The patient will be allowed to breathe normally to acclimate themselves to the mask and airflow. The patient will be allowed to hold the device, if desired, and a head strap will be placed around their head for additional support and comfort. Once the patient is acclimated (<1 min), they will be prompted to start the collection. Pre-loaded collection parameters will be set with the device software to simultaneously collect four tubes of exhaled breath from patient's expired breath. Each metal tube (pictured in Figure 1) is packed with material to bind volatile chemicals. Once collection begins, the system will learn the breathing pattern of the patient (25 sec) and will adjust the collection pumps accordingly. Breath collection will take between 5-7 min, depending on the patient's breathing rate and total collection volume. This study will collection 500 mL of exhaled breath on each of the four tubes. All exhaled breath samples will be stored at 4°C until analysis. For exhaled breath condensate, we

Figure 3: R-tube breath collection device

will utilize the R-tube collection device (Respiratory Research Inc.) (Figure 3). The R-tube is a one-time use apparatus that utilizes a metal cooling sleeve, placed over the plastic tube, to cool the patient's breath, creating condensation build up within the device. The patient will breathe through the mouth piece for 10 min. Following the collection, a metal plunger will be used to consolidate the condensate within the collection tube and the EBC sample can be transferred to a fresh sterile tube. The saliva, urine and blood tube will be transported to the research lab at room temperature and processed within two hours. All biofluids will be aliquoted, frozen and stored at -80° C until analysis.

**Surgical tumor removal:** In addition to our non-invasive collections, as part of the Advanced Tumor Bank protocol, the tumor tissue will be collected and banked. We will obtain tumor tissue samples from the Advanced Tumor Bank following processing.

**Post-surgery follow-up collection:** 24-48 hours prior to their 4-week in-clinic follow-up appointment, patients will be called to remind them for their 4-week follow-up, and follow-up collections will be performed in Dr. Wudel's clinic, 4$^{th}$ floor of the Cancer Center. Upon entering the clinic, the patient will be taken the research nurse visit room. The research nurse will obtain current medications and tobacco use information. The patient will be prompted for a urine sample. If the patient again agrees to a blood sample, the research nurse will collect 1 redtop vacutainer serum tube and 1 purpletop vacutainer EDTA tube from the patient. We will initiate the saliva and exhaled breath collection as described above. Following all of the collections, the patient will be given their incentives for participating in the study.

All samples collected from this study will be banked and will have the potential to be used for future scientific experimentation, which may include additional genetic, proteomic, and metabolomic analyses not described in the data collection section below.

**Research Data Collection:**

**Data generation:** Unbiased metabolomics analysis of bio-volatiles and biofluids has the potential to identify 100-1000's of metabolites in any single sample. Our group has successfully assessed bio-volatiles and biofluids by unbiased metabolomics (15) (16) Although there is no single method that allows for the complete analysis of the metabolite, utilizing multiple modes of sample preparation and analysis will potentially identify a number of classes of compounds relevant to this study. The bio-volatile that will potentially be identified are various hydrocarbons, aldehydes, ketones, and alcohols. Through our biofluids analysis we will potentially identify amino acids, sugars, nucleotides, citric acid cycle intermediates through our GCMS analysis while our LCMS analysis will identify a wide range of lipids (e.g. ceramides, sphingomyelins, cholesteryl esters, oxysterols, lyso- and phospholipids, mono-, di- and triacylglycerols, galactosyl- and glucuronyllipids, bile acids, steroids, eicosanoids/oxylipins, very short chain fatty acids).

Metagenomics assessment of the microbiome will identify populations of microbes that are present in patients at the time of collections. We will map sequencing reads to a reference database consisting of complete genomes for prokaryotes (bacteria and archaea), eukaryotes (fungi and protozoa) and viruses to identify changes in these populations down to the species or strain level. We will also identify potentially active functional profiles of the microbiota identified in the samples. We will perform a de novo assembly of the reads to identify protein coding regions, and the coding regions will subsequently be annotated with Pfam domains. The abundance of functional categories will then be estimated to build a functional profile for each sample

**Breath metabolomics analysis:** Exhaled breath samples will be analyzed using a Markes thermal desorber, TD100-xr, to inject breath samples onto an Agilent two-dimensional GC system 7890B in tandem with LECO's Pegasus 4D Time-of-Flight instrument system (TD-GCxGC-TOFMS). The collected spectra will be processed using LECO's ChromaTOF version 4.72 software. Initial data processing includes peak deconvolution, 1st and 2nd dimension retention time recordings, signal to noise ratio calculation, unique mass detection, peak area determination, and compound identification based on the spectral pattern similarity match against established mass spectral reference libraries (NIST14, 2014 National Institute of Standard Technology, NIH, USA). Identified spectral matches will be aligned using R-program R2DGC (17) to determine common metabolites in patients prior to and after surgery. MetaboAnalyst 4.0 will be employed to normalize and transform the data for univariate and multivariate statistical analysis (18). Additional statistical analysis may be performed using Graphpad Prism 8 (Graphpad Software, Inc)

**Biofluid and tissue metabolomic analysis:** Saliva, serum, urine and tissue aliquots will be prepared for analysis by standard processing according to published methodology(5); (19). Two platforms of unbiased metabolomics analysis will be completed using both a Q Exactive GC system and a Q Exactive LC system. Utilizing two chromatographic methodologies will maximize our metabolite identification coverage. The collected spectra will be processed using Thermo Scientific™ TraceFinder™ 4.1 software. The initial data processing will include baseline correction, peak deconvolution, retention index, signal to noise ratio calculation, unique mass detection, peak area determination, and compound identification based on the spectral pattern similarity match against established mass spectral reference libraries (NIST14, National Institute of Standard Technology, NIH and a high-resolution mass spectral library from TraceFinder™ software). Identified spectral matches will be identified for pre- and post-surgery following metabolomics standards initiative (MSI) guidelines for metabolite identification (Level 2), with a match score requirement of equal to or greater than 80%. Aligned metabolomic data analysis software, MetaboAnalyst 4.0, will be employed to normalize and transform the data for univariate and multivariate statistical analysis (18). Additional statistical analysis may be performed using Graphpad Prism 8 (Graphpad Software, Inc)

**Biofluid and tissue microbiome analysis:** DNA will be extracted from each blood, saliva and lung tumor tissue sample using the QIAamp® UCP Pathogen Mini Kit (Qiagen) with extended homogenization of the tissue samples using the Mini-Beadbeater-96 cell disrupter (Biospec Products). The quantity and quality of the DNA extracts will be determined using a DeNovix DS-11 FX+ Spectrophotometer/Fluorometer and Agilent 4200 TapeStation, respectively. The DNA will then be prepared for shotgun metagenomic sequencing using the Nextera® DNA Flex Library Preparation Kit (Illumina). DNA libraries will be normalized and quantified prior to pooling to ensure equal library representation. Pooled DNA libraries will then be sequenced using the Illumina NovaSeq 6000 S1 Reagent Kit and sequencer, generating 2 x 150 bp paired-end reads. Raw sequence reads (FASTQ files) will be imported into the Microbial Genomics Pro Suite within the CLC Genomics Workbench (version 11.0.1; https://www.qiagenbioinformatics.com) for QC, adapter trimming, and analysis. High-quality (≥Q30) reads for each sample will be aligned to those in a blank extraction sample (negative control) that will be processed from DNA extraction through library preparation to identify and remove contaminating sequences. The filtered reads will then be aligned to a reference database consisting of complete genomes in the RefSeq database for prokaryotes (bacteria and archaea), eukaryotes (fungi and protozoa) and viruses for taxonomic profiling. The filtered reads will also be assembled into contigs *de novo*, and MetaGeneMark (20) will be used to identify protein-coding regions within the reads to predict functional diversity of the microbiota identified in the samples. The coding regions will be annotated with Pfam domains

and GO terms, and the abundance of functional categories will be estimated to build a functional profile for each sample.

The results from the taxonomic and functional profiling will each be merged to estimate similarities and differences among the samples. Heat maps will be generated using Euclidean distance and clustering based on average linkage to assess functional similarities between samples, and principal component analysis (PCA) ordination of Bray-Curtis similarity between samples will be used to visualize the microbiota diversity across the samples.

**Data Storage:** All research data collected on this study will be stored on Wake Forest Baptist Medical Center password protected computers and servers.

## 4.3 Study-Related Activities

| | Pre-Study | Pre-surgery Collection | Surgical tumor removal | Post-Surgery Follow-up |
|---|---|---|---|---|
| Informed consent | X | | | |
| Demographics | X | | | |
| Medical history | X | | | |
| Pulmonary Function Tests (PFTs) | | X | | |
| Concurrent meds | X | | | X |
| Smoking Status | | X | | X |
| Breath Collection (EBC[a]/EBV[b]) | | X | | X |
| Saliva collection[c] | | X | | X |
| Blood collection[d] | | X | | X |
| Urine collection[e] | | X | | X |
| Tumor collections | | | X | |
| Adverse event evaluation | | X | | X |

[a] Approximately 1-1.5 mL of EBC will be generated in 10 min of breathing
[b] 500 ml of EBV in each of 4 collection tubes will be collected in 5-7 min of breathing
[c] 2 ml minimum of saliva will be collected into SalivaBio Passive Drool collection tubes
[d] 5-10 ml blood sample in each redtop and purple top vacutainer blood collection tube
[e] 25 ml minimum of urine will be collected in sterile specimen containers

# 5.0 Outcome Measures

## 5.1 Primary Outcomes

5.1.1 Since the primary objectives is to assess the feasibility of adding these non-invasive sample collections to pre- and post-surgery visits the following outcomes (counts) will be gathered to examine this:

5.1.1.1 Number of patients who are screened for trial

5.1.1.1.1 Among those screened, the number who are eligible
5.1.1.1.2 Among those eligible, the number who consent to participate
5.1.1.1.3 Among those who participate, the number who provide pre-surgery samples
5.1.1.1.4 Among those who provide pre-surgery samples, the number who provide post-surgery samples

## 5.2 Secondary Outcomes

5.2.1 Since the first secondary objective is to examine metabolic and microbial signatures are two time points (pre- and post- surgery) therefore the following outcomes will be gathered to examine this:

5.2.1.1 Quantitative assessments of metabolic characteristics measured at pre-surgery
5.2.1.2 Quantitative assessments of microbial assessments measured at pre-surgery
5.2.1.3 Quantitative assessments of metabolic characteristics measured at post-surgery
5.2.1.4 Quantitative assessments of microbial assessments measured at post-surgery

Using these four sets of quantitative assessments, "signatures" will be developed (described in statistical analysis section) and compared to see whether there are characteristics that remain similar or change when comparing pre- and post-surgical measures.

5.2.2 Patient LC Stage will be gathered as a secondary outcome. This outcome will be used with the outcome measures gathered for primary outcome 5.1.2 (described above) to address the secondary objective (To identify signatures which are associated with LC stage).

5.2.3 Patient pulmonary function status (assessed by Dr. Wudel) will be gathered as a secondary outcome. This outcome will be used with the outcome measures gathered for primary outcome 5.1.2 (described above) to address the secondary objective (To identify signatures which are impacted by patient's pulmonary function status).

# 6.0 Analytic Plan

## 6.1 Sample Size and Power

This is a pilot study therefore; many of the objectives are exploratory in nature. With this in mind, we plan to recruit two cohorts of patients with this study, 45 female and 45 males. In order to be able to examine the questions concerning the possible relationships of metabolic and microbial signatures over time and with lung cancer outcomes, both pre- and post- surgery measures are needed. Therefore, each patient must provide both a pre- and post-surgery sample (optional blood will be allowed) in order to be used in the signature analyses. If a post collection is not performed or collection is unsatisfactory, the patient will be replaced. Therefore, our target sample size of 90 patients (45 male/45 female) is the target for patients with both pre- and post- samples collected.

First, for the primary objective of assessing feasibility, we can determine the total number of patients needed in order to enroll 90 patients. To examine this question we will determine 1) the number of patients screened for the trial, 2) among those screened, the number who are eligible, 3) among those eligible, the number who consent to participate, 4) among those who participate, the number who provide pre-surgery samples, and 5) among those who provide pre-surgery samples, the number who provide post-surgery samples. Based on our design we will enroll patients until 90 provide pre- and post- surgery samples.

Since we do not know prospectively how many patients will be needed to enroll 90 we will conservatively estimate several possible scenarios and for each scenario describe the width of the corresponding 95% Exact Clopper Pearson confidence interval for proportions.

The table below shows different total number of patients needed to enroll 90 for the study, beginning with 120 (75% enrollment rate) to 180 (50% enrollment rate).

| # screened and eligible | Number enrolled | Enrollment rate | Lower 95% limit | Upper 95% limit |
|---|---|---|---|---|
| 120 | 90 | 75% | 0.66 | 0.82 |
| 150 | 90 | 60% | 0.52 | 0.68 |
| 180 | 90 | 50% | 0.42 | 0.58 |

As can be seen from the above, the confidence intervals for each scenario are approximately +/- 10% or less depending on the scenario observed.

For the other primary outcomes, metabolic and microbial assessments taken at two time points (pre- and post- surgery), we can determine statistical power in several ways. One straight-forward approach will be to assess the difference in biomarkers (metabolic or microbial) that can be detected when comparing pre- and post- surgical assessments within patients. If there are 90 patients then there is 80% power (using an alpha=0.05, 2-sided) to detect a difference in a biomarker measure equal to 0.30 standard deviations for that marker.

Since this is a pilot study, the above power calculations suggest that useful data concerning feasibility and effect size assessments for biomarkers can be determined in this study. Multiple comparison adjustments are not considered at this time since the focus of examining these effect size estimates is to determine an overall ranking of measures, in order to inform future studies concerning which signatures may be most promising to examine further.

## 6.2 Analysis of Primary Outcome

To examine the questions of feasibility, we will generate a series of estimates of proportions for each question (i.e. proportion of screened who are eligible, etc.). Next, we will estimate 95% exact Clopper-Pearson confidence intervals for each feasibility estimate.

## 6.3 Analysis of Secondary Outcomes

For the analysis of metabolic and microbial measures we will examine each marker assessed both at pre- and post- surgery. We hypothesize that markers that are associated with lung cancer will change between the pre- and post- surgery setting, whereas markers that are independent of lung cancer will not change between the two time points. Therefore, a first analysis will be to examine

a series of paired t-tests to determine which markers significantly change between the two time points for patients.

In addition to this analysis, we wish to determine whether there are any marker signatures that appear in groups of patients. To do this we will use cluster analysis methods to determine whether there are any groups of patients who share similar marker profiles pre-, post- or change from pre- to post- for metabolic and microbial measures. If we are able to identify clusters of patients with common marker signatures, we will then determine whether there are other disease characteristics that are shared by these patients (i.e., tumor stage, lung function, etc.). This analysis approach will also be used below to address the analysis of secondary outcomes. As stated above, this is a pilot study, therefore these cluster analyses will be considered exploratory and hypothesis generating.

To address the two secondary objectives, we will use the same data gathered for the primary objective as described above, however we will compare those data to the two secondary outcomes measures (LC Stage and patient pulmonary function). The analytic difference between these two measures is that LC stage will be an ordinal variable and pulmonary function is a continuous measure.

For LC Stage, we will examine whether pre-surgery, post-surgery or change from pre- to post-surgery metabolic or microbial measures are different depending on LC Stage. To do this we can consider the metabolic or microbial measures as outcomes and use a general linear models framework to see if LC Stage (included as a class variable, but then examined as an ordinal variable using a contrast statement in GLM) is associated with any marker.

In addition to this analysis where individual markers are considered separately, we will use cluster analyses (as described above) based on the marker data and then determine whether there are associations between cluster membership and LC Stage using 2-way contingency tables.

For pulmonary function, we will use the same approach as for LC Stage, except that pulmonary function will be considered as a continuous variable in the model. Thus, we can examine associations (correlations) between the marker measures and pulmonary function. In addition, pulmonary function can be examined at both time-points (pre- and post- surgery) thus we can examine the relationships between pulmonary function measures and biomarker measures at each time points (and the change in measures over time).

In addition to these analyses, additional exploratory analyses will be examined to determine whether other patient level characteristics (i.e., gender, age, smoking history (current/former/never), etc.) are associated with different metabolic and/or microbial signatures or measures in a general linear model framework.

Since this is a pilot study and the sample size is moderate (n=90), these additional analyses will be primarily hypothesis generating to allow us to determine whether there are potential signals in the data that may need to be further examined in larger trials.

## 6.4 Accrual Rate

Based on the large number of anticipated patients for this trial and the non-invasive nature of the measurement we anticipate that the majority of eligible patients will participate. With this in mind, we anticipate that we will enroll at least 8 patients per month (2 per week), and thus complete accrual within 60 weeks (14 months). This allows for the possibility that up to 120 patients are screened, enrolled and provide pre-surgery data in order to have 90 patients also provide post-

surgery data (i.e. 75% of patients who provide pre-surgery data also provide post-surgery data). We believe this is a conservative estimate for the accrual rate, however we recognize that the trial may take longer (or shorter) to be completed based on the actual rate of completion of patients (which is one of the questions to be answered by the trial). With this in mind the shortest anticipated time for the trial is 45 weeks (2/week x 45=90) and the longest anticipated time is 90 weeks, allowing for 50% of patients who enroll and provide pre-surgery data to not provide post-surgery data.

There are no interim analyses planned for this study.

## 6.5 Length of Study

Once all patients are enrolled the last follow-up measure will happen 1 month post-surgery, therefore we anticipate that the study will be completed within 15 months of when enrollment begins (14 months for accrual, one month for final follow-up).

# 7.0 Data Management

| | |
|---|---|
| Informed consent document | EPIC |
| Protocol registration form | WISER/OnCore |
| Medical History | WISER/OnCore |
| Specimen Collection Form | WISER/OnCore |
| Concurrent Medications Form | WISER/OnCore |
| Tobacco Use Form | WISER/OnCore |
| PFT Form | WISER/OnCore |
| Final Pathology Report Form | WISER/OnCore |
| Adverse Event Log | WISER/OnCore |
| Research Data | Secure research lab computers and servers |

# 8.0 Confidentiality and Privacy

Confidentiality will be protected by collecting only information needed to assess study outcomes, minimizing to the fullest extent possible the collection of any information that could directly identify subjects, and maintaining all study information in a secure manner. To help ensure subject privacy and confidentiality, only a unique study identifier will appear on the data collection form. Any collected patient identifying information corresponding to the unique study identifier will be maintained on a linkage file, stored separately from the data. The linkage file will be kept secure, with access limited to designated study personnel. Following data collection, subject identifying information will be destroyed after 5 years, consistent with data validation and study design, producing an anonymous analytical data set. Data access will be limited to study staff. Data and records will be kept locked and secured, with any computer data password protected. No reference to any individual participant will appear in reports, presentations, or publications that may arise from the study.

# 9.0 Data Safety and Monitoring

The principal investigator will be responsible for the overall monitoring of the data and safety of study participants. The principal investigator will be assisted by other members of the study staff.

## 10.0 WFUHS IRB AE Reporting Requirements

Any unanticipated problems involving risks to subjects or others and adverse events shall be promptly reported to the IRB, according to institutional policy. Reporting to the IRB is required regardless of the funding source, study sponsor, or whether the event involves an investigational or marketed drug, biologic or device. Reportable events are not limited to physical injury, but include psychological, economic and social harm. Reportable events may arise as a result of drugs, biological agents, devices, procedures or other interventions, or as a result of questionnaires, surveys, observations or other interactions with research subjects.

All members of the research team are responsible for the appropriate reporting to the IRB and other applicable parties of unanticipated problems involving risk to subjects or others. The Principal Investigator, however, is ultimately responsible for ensuring the prompt reporting of unanticipated problems involving risk to subjects or others to the IRB. The Principal Investigator is also responsible for ensuring that all reported unanticipated risks to subjects and others which they receive are reviewed to determine whether the report represents a change in the risks and/or benefits to study participants, and whether any changes in the informed consent, protocol or other study-related documents are required.

Any unanticipated problems involving risks to subjects or others occurring at a site where the study has been approved by the WFUHS IRB (internal events) must be reported to the WFUHS IRB within 7 calendar days of the investigator or other members of the study team becoming aware of the event.

Any unanticipated problems involving risks to subjects or others occurring at another site conducting the same study that has been approved by the WFUHS IRB (external events) must be reported to the WFUHS IRB within 7 calendar days of the investigator or other members of the study team becoming aware of the event.

Any event, incident, experience, or outcome that alters the risk versus potential benefit of the research and as a result warrants a substantive change in the research protocol or informed consent process/document in order to insure the safety, rights or welfare of research subjects.

## 11.0 Reporting of Unanticipated Problems, Adverse Events or Deviations

Any unanticipated problems, deviations or protocol changes will be promptly reported by the principal investigator or designated member of the research team to the IRB and sponsor or appropriate government agency if appropriate.

## References

1.Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7-30.
2.Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. J Thorac Oncol. 2016;11(1):39-51.
3.Dotson T, Bellinger C. Molecular Diagnostics to Enhance Lung Cancer Screening & Diagnosis. Oncology Times. 2016;38(24):30-1.
4.National Lung Screening Trial Research T, Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, et al. The National Lung Screening Trial: overview and study design. Radiology. 2011;258(1):243-53.
5.Fiehn O. Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling. Curr Protoc Mol Biol. 2016;114:30 4 1- 4 2.
6.Vasilopoulou CG, Margarity M, Klapa MI. Metabolomic Analysis in Brain Research: Opportunities and Challenges. Front Physiol. 2016;7:183.
7.Fiehn O. Metabolomics--the link between genotypes and phenotypes. Plant Mol Biol. 2002;48(1-2):155-71.
8.Bamji-Stocke S, van Berkel V, Miller DM, Frieboes HB. A review of metabolism-associated biomarkers in lung cancer diagnosis and treatment. Metabolomics. 2018;14(6).
9.Chen J, Domingue JC, Sears CL. Microbiota dysbiosis in select human cancers: Evidence of association and causality. Semin Immunol. 2017;32:25-34.
10.Huffnagle GB, Dickson RP. The bacterial microbiota in inflammatory lung diseases. Clin Immunol. 2015;159(2):177-82.
11.Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. Am J Respir Crit Care Med. 2011;184(8):957-63.
12.Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care Med. 2013;187(10):1067-75.
13.Vogtmann E, Goedert JJ. Epidemiologic studies of the human microbiome and cancer. Br J Cancer. 2016;114(3):237-42.
14.Zeng XT, Xia LY, Zhang YG, Li S, Leng WD, Kwong JS. Periodontal Disease and Incident Lung Cancer Risk: A Meta-Analysis of Cohort Studies. J Periodontol. 2016;87(10):1158-64.
15.Bishop AC, Libardoni M, Choudary A, Misra B, Lange K, Bernal J, et al. Nonhuman primate breath volatile organic compounds associate with developmental programming and cardio-metabolic status. J Breath Res. 2018;12(3):036016.
16.Misra BB, Bassey E, Bishop AC, Kusel DT, Cox LA, Olivier M. High Resolution GC/MS Metabolomics of Non-Human Primate Serum. Rapid Commun Mass Spectrom. 2018.
17.Ramaker RC, Gordon ER, Cooper SJ. R2DGC: threshold-free peak alignment and identification for 2D gas chromatography-mass spectrometry in R. Bioinformatics. 2018;34(10):1789-91.
18.Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res. 2018;46(W1):W486-W94.
19.Misra BB, Upadhayay, R.P., Cox, L.A.; Olivier, M. Optimized GC–MS metabolomics for the analysis of kidney tissue metabolites. Metabolomics. 2018;14(75).
20.Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 2010;38(12):e132.

# Appendix A – Subject Eligibility Checklist

| **IRB Protocol No.** | **WFBCCC Protocol No. 03219** |
|---|---|
| **Study Title:** Pilot Study: Metabolic and microbial profiling of lung cancer | |
| **Principal Investigator: Andrew Bishop, Ph.D.** | |

| **Inclusion Criteria (as outlined in study protocol)** | **Criteria is met** | **Criteria is NOT met** | **Source Used to Confirm * (Please document dates and lab results)** |
|---|---|---|---|
| Male and female patients age ≥18 years, of all racial and ethnic origins, with suspected non-small cell lung cancer stages I, II, and III as evident through radiographic evidence and felt acceptable to undergo surgical resection. | ☐ | ☐ | |
| Patients who have the ability to understand and the willingness to sign a written consent form. | ☐ | ☐ | |
| **Exclusion Criteria (as outlined in study protocol)** | **Criteria NOT present** | **Criteria is present** | **Source Used to Confirm * (Please document dates and lab results)** |
| Patients who are have taken antibiotics within two weeks | ☐ | ☐ | |
| Patients currently undergoing active treatment for other malignancies | ☐ | ☐ | |
| Subjects who are unable or unwilling to provide consent | ☐ | ☐ | |
| Patients who are on continuous supplemental oxygen | ☐ | ☐ | |

This subject is ☐ eligible / ☐ ineligible for participation in this study.

OnCore Assigned PID: ___ ___ ___ ___ ___ ___ ___ ___

Signature of research professional confirming eligibility:_____
Date: ___ ___/___ ___/___ ___

Signature of Treating Physician**: _____
Date: ___ ___/___ ___/___ ___

* Examples of source documents include clinic note, pathology report, laboratory results, etc. When listing the source, specifically state which document in the medical record was used to assess eligibility. Also include the date on the document. Example: “Pathology report, 01/01/14” or “Clinic note, 01/01/14”

**Principal Investigator signature can be obtained following registration if needed

# Appendix B – Protocol Registration Form

DEMOGRAPHICS

Patient: Last Name: _____ First Name: _____

MRN: ___ ___ ___ ___ ___ ___ ___ ___ DOB (mm/dd/yy): ___ ___ / ___ ___ / ___ ___

ZIPCODE: ___ ___ ___ ___ ___

SEX: ☐ Male ☐ Female Ethnicity (choose one): ☐ Hispanic ☐Non-Hispanic

Race (choose all that apply): ☐ WHITE ☐BLACK ☐ ASIAN ☐ PACIFIC ISLANDER ☐ NATIVE AMERICAN

Height: ___ ___.___ inches Weight: ___ ___ ___.___lbs.(actual)

Surface Area: ___ ___.___$m^2$

Primary Diagnosis: _____

Date of Diagnosis: ___ ___ /___ ___/___ ___

Performance Status: ___ ☐ ECOG

PROTOCOL INFORMATION

Date of Registration: ___ ___ / ___ ___ / ___ ___

MD Name (last) : _____

Informed written consent: ☐ YES ☐ NO

(consent must be signed prior to registration)

Date Consent Signed: ___ ___ / ___ ___ / ___ ___

PID # (to be assigned by OnCore): ___ ___ ___ ___ ___ ___ ___ ___

*Protocol Registrar can be contact by calling 336-713-6767 between 8:30 AM and 4:00 PM, Monday – Friday.*

*Completed Eligibility Checklist and Protocol Registration Form must be hand delivered, faxed or e-mailed to the registrar at 336-7136772 or registra@wakehealth.edu.*