Integrative Modeling of Multi-Omic Data Using a Mediation Framework

Miranda Kroehl, PhD
Department of Biostatistics and Informatics
Colorado School of Public Health
University of Colorado, Anschutz Medical Campus

Daniel N. Frank, PhD
Division of Infectious Disease
School of Medicine
University of Colorado, Anschutz Medical Campus

CCTSI Cores Utilized:
Biostatistics, Epidemiology, & Research Design (BERD)

2. Specific Aims

In 2017, an estimated 65,000 new diagnoses were made of head and neck squamous cell carcinomas (HNSCC) in the United States[1]. Despite concerted efforts to improve diagnosis and treatment, HNSCCs remain lethal with 5-year survival rates as low as 33%. Emerging evidence suggests that changes in the oral microbiome are associated with HNSCC[2–4]. Furthermore, bacteria have been linked to carcinogenesis through mechanisms such as the production of toxic metabolites[5]. In order to better understand the role of the microbiome in the initiation, progression, and treatment of HNSCC and cancer in general, we and others are employing a variety of 'Omics technologies to identify the critical pathway(s) and key microbiome-host interactions that shape carcinogenesis. Through next generation sequencing, mass spectrometry, and similar technologies, it is now straightforward to interrogate specimens to this end. However, progress is hindered due to a lack of adequate statistical tools which can distill a large number of predictors, mediators, and covariates and analyze these multi-'Omic datasets in a comprehensive fashion. **Therefore, the goal of this project is to develop robust statistical methods for integrative analysis of multiple 'Omics datasets** that are applicable not only to our current HNSCC study, but more-widely translatable to studies of microbiome-host interactions in all fields. This proposal lays out and investigates approaches to solve this methodological gap in the following aims:

Specific Aim 1: Develop and test a global intersection union test (gIUT) for modeling microbiome, metabolome, and outcome data under a mediation framework. Using an established framework for mediation analysis often referred to as the causal steps approach, and incorporating regression approaches that can handle high-dimensional predictors or mediators that will allow one to test for associations between multi 'Omics datasets and study outcomes. We hypothesize that the proposed statistical test will maintain type I error rates at nominal levels and have reasonable power, and will demonstrate this through simulation studies and apply this method to existing data for our ongoing HNSCC study.

Specific Aim 2: Develop and test multiple estimators for the indirect effect of the microbiome – metabolome – outcome relationship using a mediation framework. A second approach for assessing mediation is by estimating the indirect effect of the exposure on outcome, acting through the mediating variable. We propose two separate estimators for the indirect effect of the microbiome on outcomes: (1) a global estimate for the indirect effect (GIE) and (2) taxon-specific estimates for the indirect effect (TIE), and corresponding statistical tests. We hypothesize the proposed tests will maintain reasonable statistical properties and will, in some circumstances, be more powerful than the gIUT proposed in Aim 1. We will demonstrate this through simulation studies and apply these methods to existing data for our ongoing HNSCC study.

3. Background and Significance

The Microbiome and HNSCC. An estimated 30% of all rare cancers diagnosed in 2017, including those of the oral and nasal cavities, pharynx, and larynx, are classified as HNSCC[1]. Although some environmental toxins have been linked to HNSCC, current knowledge is insufficient to understand development and progression of disease or response to therapeutics. Individual bacteria species and alterations in the composition of the human microbiome (dysbiosis) have been associated with many cancers[7–13], including HNSCC[2–4]. Furthermore, the microbiome influences several key hallmarks in cancer development, such as metabolism and anti-cancer immunity[5,14]. In patients with HNSCC, carcinogenic metabolites have been found in the oral cavity, but exactly where and how they are generated has not yet been explored[15].

To better understand the relationships between the oral microbiome, metabolome, and HNSCC, we have an established cross-sectional cohort of 72 HNSCC patients and 83 healthy controls from whom saliva samples had been obtained (COMIRB #12-1328 and #16-1794) and 16S rRNA gene sequencing was applied. We hypothesize that changes in the balance between cancer-promoting and protective microbes contribute to cancer development, and have identified community-wide differences between HNSCC cases and controls at the phyla and genera levels ($p=0.0008$ and $p<0.0001$ respectively) using a permutation based analysis of variance[16] (PERMANOVA). Furthermore, numerous individual bacterial genera and phyla also differed in relative abundance (RA) between the groups (Fig. 1), even after accounting for environmental exposures. Although we have identified differences between the oral microbiome and HNSCC, we have yet to determine how the metabolome mediates this association.

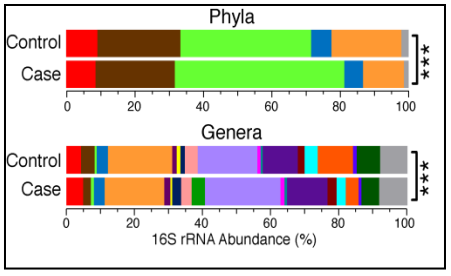

Fig. 1. Variation in Oral Microbiome between HNSCC Cases and Controls The bar chart summarizes the mean RA of predominant bacterial phyla and genera, highlighting differences between HNSCC patients and healthy controls.

Key Methodological Gaps. The pathways through which microbes impact disease remain to be elucidated in detail and validated[17]. One barrier to this knowledge comes from statistical challenges with data integration between the microbiome and other 'Omics technologies, such as metabolomics. Data integration, the process of combining and interrogating data from multiple sources, can be accomplished using a variety of techniques ranging from simple univariate or multivariate correlations (e.g. Spearman correlation, principal components analysis) to more sophisticated network-based approaches[18–22]. Often, integration requires data reduction methods, such as networks (e.g. co-abundance networks for microbiota) due to the high-dimensional nature of 'Omics data, and consequently these approaches do not allow for interpretation of the individual components within a dataset. *Of note, most integration techniques are limited to describing associations between two 'Omics datasets and do not evaluate subsequent associations with outcomes*.

Microbiome data are particularly challenging to work with because they are compositional, non-negative, and zero-inflated, and consequently many standard integration techniques are not appropriate[23,24]. Recent literature stresses the need for improved statistical methods, especially data integration techniques that model relationships between the microbiome and other types of high-dimensional data[19,24]. A causal mediation framework is one approach that is used to evaluate how an outcome is affected through an exposure variable[25–27]. Our prior work extends upon traditional mediation models to evaluate mediation with a set of correlated high-dimensional exposure variables, such as microbiota[28]. This framework has also been extended to accommodate a set of high-dimensional mediators, such as gene expression data[29,30], yet no methods exist to evaluate both multiple exposures and multiple mediators.

In this proposal, we describe two variations on a method for data integration of both multiple exposures and multiple mediators under a mediation framework. Our methods will permit modeling in a high-dimensional setting and thus examination of the individual components (e.g. individual taxa in a microbial community, individual chemical compounds in a metabolome), as well as their interrelationships with each other and with clinical outcomes. We believe this approach will provide several novel advantages, such as additional information gained and identification of specific components to target for future interventions, compared with other commonly used microbiome data-reduction techniques that involve simple exploratory graphical analysis (e.g., PCA/PCoA) or reduction of data to a single values using co-abundance and co-expression networks and inserting these values into a mediation analysis. As a test case, we will apply our methods to an existing HNSCC dataset in order to gain a better understanding of the microbiome, metabolites, cancer phenotype in order to close the existing knowledge gap, allowing for better prevention and treatment plans. Furthermore, the approaches we are developing are not limited to HNSCC, and therefore will be highly translatable to other areas of research. *Completion of this study will generate novel analytic methods and critical preliminary data, help refine hypotheses, and lead to joint publications, all of which will increase the competitiveness of our planned grant proposals;* for instance, for the HNSCC project, we anticipate submitting both R01 and SPORE grants in 2018. Other funded projects unrelated to HNSCC (see Other Support, below) as well as planned R01 proposals also will benefit from development of these methods.

4. Novel Methodology

Overview: The long-term goal of this work is to provide an analytic framework for integrating multiple high-dimensional datasets and quantifying their impact on health outcomes. The central hypothesis of this proposal is that a causal mediation framework may be used to uncover new information on how metabolites mediate the relationship between the microbiome and development of disease (using HNSCC as a motivating example). To test this hypothesis, we will develop a novel integrative modeling approach of the microbiome and metabolomics data and evaluate an intersection-union based hypothesis test of individual regression coefficients (Aim 1) and a hypothesis test of both global and taxa-specific indirect effect estimates (Aim 2). Future directions for this work include application to other disease states, other 'Omics types, and extensions to accommodate for interventions acting upon the microbiome.

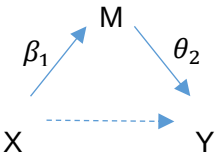

Fig. 2. Mediation. The following two regression models:

$E[Y|X,M] = \theta_0 + \theta_1 X + \theta_2 M$

$E[M|X] = \beta_0 + \beta_1 X$

can be used to estimate the indirect effect of X on Y, acting through M. The natural indirect effect is defined as $NIE = \beta_1\theta_2$.

Specific Aim 1: Develop and test a global intersection union test (gIUT) for modeling microbiome, metabolome, and outcome data under a mediation framework.

Mediation analysis can be used to investigate the mechanisms through which an exposure, X, affects an outcome, Y, through an intermediary variable, M. The mediating variable is on the causal pathway connecting the two; this pathway,

portrayed by the solid arrows (Fig. 2), is referred to as the indirect effect of X on Y. The direct effect, represented by the dashed arrow (Fig. 2), is the effect of X on Y, not acting through the intermediate variable. One approach for testing mediation is the *causal steps* approach, common in psychological and behavioral research fields[31–33]. Under this approach, the two regression models for mediation (Fig. 2) are fit, and the individual coefficients along the indirect effect pathway are tested. This dual hypothesis testing can be conducted as an intersection-union test[34], in which the null hypothesis of no mediation will be rejected only if both individual hypotheses are rejected. We note here that the arguments for mediation, and especially causality, are not limited to results from a statistical test, and thus in order for one to reasonably conclude that mediation exists, the design and conceptual assumptions for mediation must be met in conjunction with a statistically significant finding.

To evaluate mediation with both high-dimensional exposures and high-dimensional mediators, we posit the relationship between Y, a set of mediators M, and a set of predictors X, can be modeled as:

$$E\left[Y\middle|X_1,\dots,X_p,M_1,\dots,M_q\right]=\theta_0+\sum_{j=1}^{q}\theta_j M_{\boldsymbol{m}}+\sum_{i=1}^{p}\gamma_i X_i \quad (1)$$

$$\sum_{j=1}^{q}\theta_j M_m=\beta_0+\sum_{i=1}^{p}\beta_1 X_i+\epsilon_i \quad (2)$$

where p is the number of observed microbial taxa and q is the number of observed metabolites. For simplicity, we present the scenario where Y and M are continuous and normally distributed, and there are no covariates or exposure-mediator interactions, however, this framework allows for non-normal outcomes, as well as covariates and interactions[25,29,30]. Using the same theoretical assumptions as Zhao et al. (2014), equation 1 will be estimated under the assumption that $\gamma_i=0$ (i.e. no direct effect of X_i on Y). Both equations 1 and 2 will be estimated using lasso regression, a penalized regression approach appropriate for high-dimensionality[35–37].

The global intersection-union test for integrative microbiome analysis (gIUT), builds upon the "causal steps" approaches for testing mediation. The premise for this test is that the pathways between X-M and M-Y will be evaluated separately under the null:

$$H_{0,gIUT}\colon\{H_{0,1}\colon\theta_1=\cdots=\boldsymbol{q}=0, or\ H_{0,2}\colon\beta_1=\cdots=\boldsymbol{\beta_p}\} \quad (3)$$

The *pseudo-F* test for the LASSO[28], a novel permutation-based approximation to a usual *F* test for evaluation of coefficients developed by Dr. Kroehl, will be used to evaluate each component of the gIUT null. The gIUT hypothesis may only be rejected if both $H_{0,1}$ and $H_{0,2}$ are rejected. Simulation studies are described below.

Specific Aim 2: Develop and test multiple estimators for the indirect effect of the microbiome – metabolome – outcome relationship using a mediation framework.

A second approach for evaluating mediation is to estimate and test the indirect effect. Under the causal inference approach to mediation analysis, the counterfactual definitions of direct and indirect effect effects may be estimated from regression parameters obtained from in two equations[25–27]. The natural indirect effect (NIE), quantifies the relationship of X on Y through mediator M (Fig. 2).

Here we propose a novel estimate for the indirect effect, the global indirect effect (GIE), and corresponding hypothesis test the GIE, conducted under the null:

$$H_{0,GIE}\colon\sum_{j=1}^{q}\theta_j*\sum_{i=1}^{p}\beta_i=0 \quad (4)$$

The distribution of the GIE will be approximated and a confidence interval constructed using bootstrapping. The null hypothesis will be rejected if 0 is not contained within the confidence bounds[38].

In addition to the global tests for microbiome-metabolome mediation, we propose taxa-specific estimates of the indirect effect (TIE), as an extension of indirect effect estimates for mediation with high-dimensional continuous mediators proposed by Zhao et al. (2014) and modified by Huang and Pan (2016):

$$TIE=\sum_{j=1}^{q}\hat{\theta}_j*\hat{\beta}_i \quad (5)$$

The distributions of the TIE will be approximated using bootstrapping to obtain confidence bounds. Hypothesis testing of TIE may be conducted by specifying the null hypothesis of the TIE (e.g. $H_{0,TIE}\colon\sum_{j=1}^{q}\hat{\theta}_j*\hat{\beta}_i=0$) and evaluating using the bootstrapped confidence bounds. Simulation studies are described below.

Simulation Study for Aims 1 and 2: A simulation study will be used to empirically evaluate the behavior of the gIUT and GIE tests in terms of type I error and power, as well as tests for the TIE. We will consider scenarios with continuous and binary outcomes, sample sizes of 100, 250, and 500, and p=100, q=200. Microbiota data will be generated using the R package, MiSPU[39,40], and associations between individual microbiota and a set of

metabolites, and metabolites and Y, will be simulated as described in Zhao et. al. (2014). The null hypothesis for mediation analysis is truly a composite null hypothesis consisting of three scenarios where there is no X-M relationship, no M-Y relationship, or neither X-M nor M-Y relationships[41], and our simulations will consider each of the possible scenarios, as well as a range of alternative conditions. Under settings where an X-M or M-Y relationship is simulated, we will simulate coefficients to be non-zero with probabilities ranging from 0.1 - 0.5, and non-zero coefficients will be generated uniformly from $[-1, -0.05] \cup [0.05, 1]$. Empirical Type I error rates will be calculated as the proportion of times a test is incorrectly rejected under the null. Power will be calculated as the proportion of times a test is correctly rejected under the alternative. We will compare the performance of the gIUT test with performance of the GIE test and provide recommendations on use. It can be demonstrated that the TIE estimate is equivalent to the estimates for a single exposure and multiple mediators proposed by Huang and Pan (2016); our purpose for evaluation in this simulation is to evaluate the performance of the bootstrap test under scenarios specific to the microbiome-metabolome relationship.

Application to HNSCC data: We will apply the gIUT and GIE tests for mediation in our existing data on HNSCC patients. Saliva samples, demographics, and environmental exposure variables (smoking, drinking, and HPV) were obtained on a cohort of 72 treatment-naïve HNSCC patients and 83 healthy controls. Genomic DNAs were extracted from these saliva specimens, 16S rRNA gene PCR amplicons were generated using pan-bacterial 16S rRNA gene primers, and high-throughput sequencing was performed on an Illumina MiSeq platform operated by Dr. Frank's lab. Untargeted metabolomics analysis has been completed for 20 subjects (10 cases, 10 controls) and is underway for remaining specimens. We will model equation (1) using the binary outcome for HNSCC. Both equations (1) and (2) will be adjusted for environmental exposures (e.g., smoking/drinking/HPV and other potential confounders. If the GIE test for mediation is rejected, the TIE will be used to identify individual taxa which are relevant in the microbiome-metabolome-HNSCC pathways and may prove useful for possible interventions or improving therapeutic responses in future studies.

Possible problems and alternatives: We recognize the possibility of inconsistencies between the microbial taxa selected to have non-zero coefficients in equations (1) and (2). If these inconsistencies are problematic, we will consider alternative penalized regression approaches such as ridge regression, and L2 regularization method, or elastic net, a combination of L1 and L2 regularization. If performance of both the gIUT and GIE prove to be inadequate for hypothesis testing, we will evaluate the HNSCC using the TIE, a method with demonstrated success using SNP and gene expression data.

Timeline: We will spend months 1 and 2 developing and testing functions for our simulation studies. In month 3, we will pilot a small round of simulations and in months 4-6 we will complete simulations and analyze our findings; Aim 1 and 2 simulations will be run co-currently. In months 7-8 we will apply findings to our HNSS data, and focus on dissemination through papers, presentations, and developing a BERD workshop in months 9-12.

5. Educational Mission

Our long-term goal is to facilitate the clinical/translational applications of microbiome research by gaining a deeper understanding of how the microbiome impacts host pathophysiological processes (assessed by functional 'Omics) and subsequent health outcomes. Completion of the Aims will advance both the statistical and microbiome fields and lead to improved techniques for integrating the microbiome with metabolomics data, as well as setting the stage for integration with other 'Omics datasets. This work will have a broad impact on the greater scientific community through the development of detailed examples, tutorials, and recommendations for applying the proposed method in practice. Future efforts will be directed towards development of an open source software package for others to use.

An additional goal of this proposal is to introduce the field of integrative 'Omics research to a graduate student in Biostatistics and provide training opportunities to learn in this field. We propose to include in this project a 1st year Ph.D. Biostatistics student, Charlie Carpenter, to run data simulations and to apply the proposed methods on the oral cancer dataset. Mr. Carpenter is currently working closely under the mentorship of Drs. Kroehl, Frank, and J. Friedman and will actively participate in the writing of software, papers, and presentations related to the findings of this proposal.

The PI, Dr. Kroehl, is an active member of the Biostatistics, Epidemiology, & Research Design (BERD) program and provides education on statistical analysis of microbiome data through one-on-one grant development and BERD workshops. The findings from this work will be disseminated through a BERD open forum or workshop on how to integrate and analyze microbiome and metabolomics data. Furthermore, the methods developed in this proposal will be used to strengthen future microbiome and integrative 'Omics proposals by providing researchers with cutting edge analytics to better answer proposed research questions.

References

1. American Cancer Society. Cancer Facts & Figures 2017. Atlanta: American Cancer Society; 2017.
2. Hooper, S. J. *et al.* Viable bacteria present within oral squamous cell carcinoma tissue. *J. Clin. Microbiol.* 44, 1719–1725 (2006).
3. Guerrero-Preston, R. *et al.* 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment. *Oncotarget* 7, 51320–51334 (2016).
4. Schmidt, B. L. *et al.* Changes in Abundance of Oral Microbiota Associated with Oral Cancer. *PLoS ONE* 9, e98741 (2014).
5. Belcheva, A. *et al.* Gut Microbial Metabolism Drives Transformation of Msh2-Deficient Colon Epithelial Cells. *Cell* 158, 288–299 (2014).
6. Komatsubara, K. M. & Carvajal, R. D. The promise and challenges of rare cancer research. *Lancet Oncol.* 17, 136–138 (2016).
7. Bultman, S. J. Emerging roles of the microbiome in cancer. *Carcinogenesis* 35, 249–255 (2014).
8. Vogtmann, E. & Goedert, J. J. Epidemiologic studies of the human microbiome and cancer. *Br. J. Cancer* 114, 237–242 (2016).
9. Sun, J. & Kato, I. Gut microbiota, inflammation and colorectal cancer. *Genes Dis.* 3, 130–143 (2016).
10. Alfano, M. *et al.* The interplay of extracellular matrix and microbiome in urothelial bladder cancer. *Nat. Rev. Urol.* 13, 77–90 (2016).
11. Bhatt, A. P., Redinbo, M. R. & Bultman, S. J. The role of the microbiome in cancer development and therapy: Microbiome and Cancer. *CA. Cancer J. Clin.* 67, 326–344 (2017).
12. Mima, K. *et al.* Fusobacterium nucleatum in Colorectal Carcinoma Tissue According to Tumor Location. *Clin. Transl. Gastroenterol.* 7, e200 (2016).
13. Polk, D. B. & Peek, R. M. Helicobacter pylori: gastric cancer and beyond. *Nat. Rev. Cancer* 10, 403–414 (2010).
14. Mukherjee, P. K. *et al.* Metabolomic analysis identifies differentially produced oral metabolites, including the oncometabolite 2-hydroxyglutarate, in patients with head and neck squamous cell carcinoma. *BBA Clin.* 7, 8–15 (2017).
15. Homann, N. *et al.* Increased salivary acetaldehyde levels in heavy drinkers and smokers: a microbiological approach to oral cavity cancer. *Carcinogenesis* 21, 663–668 (2000).
16. Anderson, M. J. A new method for non-parametric multivariate analysis of variance: NON-PARAMETRIC MANOVA FOR ECOLOGY. *Austral Ecol.* 26, 32–46 (2001).
17. Frank, D. N., Zhu, W., Sartor, R. B. & Li, E. Investigating the biological and clinical significance of human dysbioses. *Trends Microbiol.* 19, 427–434 (2011).
18. Tseng, G. *Integrating omics data*. (Cambridge University Press, 2015).
19. Chong, J. & Xia, J. Computational Approaches for Integrative Analysis of the Metabolome and Microbiome. *Metabolites* 7, (2017).
20. Connor, N., Barberán, A. & Clauset, A. Using null models to infer microbial co-occurrence networks. *PLOS ONE* 12, e0176751 (2017).
21. Shaffer, M., Armstrong, A. J. S., Phelan, V. V., Reisdorph, N. & Lozupone, C. A. Microbiome and metabolome data integration provides insight into health and disease. *Transl. Res.* 189, 51–64 (2017).
22. Zhang, T. *et al.* Host Genes Related to Paneth Cells and Xenobiotic Metabolism Are Associated with Shifts in Human Ileum-Associated Microbial Composition. *PLoS ONE* 7, e30044 (2012).
23. The NIH HMP Working Group *et al.* The NIH Human Microbiome Project. *Genome Res.* 19, 2317–2323 (2009).
24. Li, H. Microbiome, Metagenomics, and High-Dimensional Compositional Data Analysis. *Annu. Rev. Stat. Its Appl.* 2, 73–94 (2015).
25. VanderWeele, T. J. A unification of mediation and interaction: a 4-way decomposition. *Epidemiol. Camb. Mass* 25, 749–761 (2014).
26. VanderWeele, T, V., S. Conceptual issues concerning mediation, interventions and composition. *Stat. Interface* 2, 457–468 (2009).
27. Valeri, L. & Vanderweele, T. J. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol. Methods* 18, 137–150 (2013).
28. Kroehl, M. On the Use of LASSO Regression for Mediation Analysis with Application to Microbiota Data. (University of Colorado Anschutz Medical Campus, 2014).

29. Huang, Y.-T. & Pan, W.-C. Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators: Hypothesis Test of Mediation Effect in Causal Mediation Model with High-Dimensional Continuous Mediators. *Biometrics* 72, 402–413 (2016).
30. Zhao, S. D., Cai, T. T. & Li, H. More powerful genetic association testing via a new statistical framework for integrative genomics: Integrative Association Testing. *Biometrics* 70, 881–890 (2014).
31. Baron, R. M. & Kenny, D. A. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182 (1986).
32. MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G. & Sheets, V. A comparison of methods to test mediation and other intervening variable effects. *Psychol. Methods* 7, 83–104 (2002).
33. Mackinnon, D. P. & Fairchild, A. J. Current Directions in Mediation Analysis. *Curr. Dir. Psychol. Sci.* 18, 16 (2009).
34. Casella, G. & Berger, R. L. *Statistical inference*. (Thomson Learning, 2002).
35. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc.* 58, 267–288 (1996).
36. Hastie, T., Tibshirani, R. & Friedman, J. H. *The elements of statistical learning data mining, inference, and prediction*. (Springer, 2009).
37. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33, (2010).
38. Manly, B. F. *Randomization and Monte Carlo methods in biology*. (Chapman & Hall, 1997).
39. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2013).
40. Wu, C., Chen, J., Kim, J. & Pan, W. An adaptive association test for microbiome data. *Genome Med.* 8, (2016).
41. Wagner, B. D. *et al.* A Multivariate Generalized Linear Model Approach to Mediation Analysis and Application of Confidence Ellipses. *Stat. Biosci.* (2017). doi:10.1007/s12561-017-9191-2

6. Budget

a. Justification.

Project Personnel

Miranda Kroehl, Ph.D., Principal Investigator (1.2 calendar months, no salary required):

Dr. Kroehl will lead this research and be responsible for all aspects of this project. She will meet with Mr. Carpenter on a weekly basis to provide guidance on coding and analytic challenges, steer the direction of the research, and trouble shoot any problems. She will meet with Dr. Frank monthly to review progress and discuss challenges and solutions as they arise.

Dan Frank, Ph.D., Co-Investigator (1.2 calendar months; no salary required):

Dr. Frank is an Associate Professor of Medicine in the Division of Infectious Diseases at the University of Colorado and a recognized expert in studies of the human microbiome in health and diseases. He will contribute microbiome and metabolome data, provide expertise in microbiome analysis, help interpret results, and contribute to software development as needed.

Charlie Carpenter, Graduate Student Research Assistant (6 calendar months):

Mr. Carpenter is currently a 1st year PhD student in the Department of Biostatistics and Informatics of the Colorado School of Public Health and University of Colorado Anschutz Medical Campus. He currently works under Dr. Kroehl to conduct analysis for microbiome studies. On this project, he will be responsible for writing code to run simulations and evaluate the proposed statistical tests, and apply these tests to the HNSCC data.

We are requesting no funds for other direct costs on this project.

b. Other support.

Other Sources of Funding

Dr. Kroehl is funded on the Biostatistics, Epidemiology, & Research Design (BERD). With this funding, she provides statistical support for grant development, educational workshops on microbiome data analysis, and conducts some methodologic research. This support will provide the unsalaried effort for Dr. Kroehl to lead this project and supervise Mr. Carpenter; this support does not provide enough effort to conduct the large amount of work needed in this proposal, hence the funding for student assistance.

Dr. Kroehl, Dr. Frank, and Mr. Carpenter are currently funded on a project with Dr. Jed Freidman to investigate the influence of maternal obesity, weight gain, and diet on the infant microbiota and programming of non-alcoholic fatty liver disease. Dr. Kroehl and Dr. Frank are also funded on a project with Dr. Adriana Weinberg examining the microbiome in HIV-exposed and unexposed infants. In both projects, the scope of work is limited to analysis of microbiome and other datasets under existing statistical frameworks and does not include the development of new statistical methodologies, as proposed here.

Microbiome and metabolomic analyses of human saliva samples from HNSCC and control patients has been supported by pilot grants from the University of Colorado Cancer Center to Dr. Frank, Dr. Shi-long Lu, and Dr. Vijay Ramakrishnan.

7. IRB, IACUC, Biosafety, and Radiation Safety approval

COMIRB #12-1328 and #16-1794

8. NIH Biosketches for each key personnel on the project

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Kroehl, Miranda Elaine

eRA COMMONS USER NAME (credential, e.g., agency login): MIRANDA.KROEHL

POSITION TITLE: Research Instructor

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Colorado State University, Fort Collins, CO	B.S	05/05	Chemical Engineering
University of Colorado Health Sciences Center, Aurora, CO	M.S	12/09	Biostatistics
University of Colorado Anschutz Medical Campus, Colorado School of Public Health, Aurora, CO	Ph.D.	05/14	Biostatistics

A. Personal Statement

As the principal investigator (PI) of this proposal, entitled **"Integrative Modeling of Multi-Omic Data Using a Mediation Framework",** I will be responsible for the development of novel statistical methodology for integrative analysis of multiple 'Omics (e.g. microbiome, metabolome, transcriptome) datasets with study outcomes using a mediation framework. I am a junior faculty member (Research Associate) in the Department of Biostatistics and Informatics of the Colorado School of Public Health. As a graduate student, I developed a novel approach to mediation analysis using LASSO regression, and proposed corresponding statistical tests appropriate for correlated high-dimensional data, such as microbiome data. My recent methodological work in mediation includes an approach for analysis using a multivariate framework, as well as work on a statistical package to evaluate unmeasured confounding, an important aspect of mediation analysis. My current collaborative work is focused on the design and analysis of microbiome-related studies, where I have been funded on NIH and industry-sponsored grants. In addition, I have worked on several microbiome data analysis projects though the Colorado Biostatistics Consortium (CBC) and Biostatistics, Epidemiology, & Research Design (BERD), where I have been funded as a statistical consultant and to provide biostatistics support for grant proposals on campus. I have worked closely with Dr. Frank over the past 3 years, and we have an excellent working relationship which has resulted in successful publications. Furthermore, I am experienced with supervising graduate research assistants and have done so for the past 4 years. My current relationship with Mr. Carpenter has, in a few short months, been successful with one submitted abstract and a paper underway. In summary, my statistical methodological and collaborative work leave me well prepared to develop the statistical methodologies proposed and result in a successful project.

B. Positions and Honors

Positions and Employment

2007-2008 Professional Research Associate, Cancer Center, University of Colorado Denver, Denver, CO
2008-2010 Graduate Research Associate, Colorado Biostatistics Consortium, University of Colorado Denver, Denver, CO
2010 Teaching Assistant, Department of Biostatistics and Informatics, University of Colorado Health Sciences Center, Denver, CO
2010-2011 Fellow, GK-12 Transforming Experiences Program, University of Colorado Denver, Denver, CO
2011-2013 Graduate Research Associate, Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO

2013-2014 Graduate Research Associate, Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO
2014-2017 Research Instructor, Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO
2017- Research Associate, Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO

Other Experience and Professional Memberships
2002-2010 Member, American Institute of Chemical Engineers
2005-2005 Member, Biomedical Engineering Society
2007- Member, American Statistical Association
2014- Member, International Biometric Society

Honors
2009 Marvin Porter Award for Outstanding M.S. Student, Department of Biostatistics and Informatics, University of Colorado Denver, Denver, CO
2010 National Science Foundation GK-12 Fellow, University of Colorado Denver, Denver, CO
2012 Orleans Research Scholarship, University of Colorado Anschutz Medical Campus, Aurora, CO
2013 1st place in Statistical Significance Poster Competition, Joint Statistical Meeting, Montreal, Quebec, Canada
2014 Strother Walker Award for Outstanding Ph.D. Student, Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO

C. Contributions to Science

1. Statistical analysis of and methods for high-dimensional data. My dissertation research involved the extension of the statistical / machine learning application, LASSO regression. I developed a method for conducting hypothesis testing and making statistical inference on coefficients obtained with these models, which are suitable for high-dimensional and highly correlated data. I then married this framework with causal mediation modeling, to investigate the role of the lung microbiome in infection and bronchiectasis in patients with cystic fibrosis (CF). In addition to statistical methods development, I have been involved in the design and analysis of many high-dimensional omics studies, including several proteomics and microbiome studies. As a collaborator with the Children's Hospital Colorado (CHCO) Breathing Institute, I have been working to help identify circulating protein biomarkers which may guide providers in the treatment of patients with cystic fibrosis (CF). Our current studies are focused on identifying low-abundance proteins, measured using SOMAScan™ technology, which are predictive of damage to lung structure (bronchiectasis) and decline in lung function. Our work was selected as the Best Abstract in Cyctic Fibrosis at the European Respiratory Society International Congress 2014, for our abstract, "Novel protein biomarkers of bronchiectasis in children with cystic fibrosis", and we have published follow-up and confirmatory studies with this information. Finally, I am also working with investigators across campus to investigate the microbiome's role in disease pathways from Type I and gestational diabetes, to immune dysregulation following HIV, malaria, and tuberculosis infection.

1. Needell, J., Dinarello, C., Ir, D., Robertson, C., Ryan, S., **Kroehl, M.**, **Frank, D.**, Zipris, D. (2017). Implication of the intestinal microbiome as a potential surrogate marker of immune responsiveness to experimental therapies in autoimmune diabetes. *PLoS One.* PubMed PMID 28301545.
2. Needell JC, Ir D, Robertson CE, **Kroehl ME, Frank DN**, Zipris D. Maternal treatment with short-chain fatty acids modulates the intestinal microbiota and immunity and ameliorates type 1 diabetes in the offspring. *PLoS One.* 2017;12(9):e0183786.
3. DeBoer, E. M., **Kroehl, M.**, Wagner, B. D., Accurso, F., Harris, J. K., Lynch, D. A., ... & Deterding, R. R. (2017). Proteomic profiling identifies novel circulating markers associated with bronchiectasis in cystic fibrosis. *Proteomics. Clinical applications*.
4. Clendenen, N., Tollefson, A., Dzieciatkowska, M., Cambiaghi, A., Ferrario, M., **Kroehl, M**., & Weitzel, N. (2017). Correlation of pre-operative plasma protein concentrations in cardiac surgery patients with bleeding outcomes using a targeted quantitative proteomics approach. *Proteomics. Clinical applications*.

2. Advancement of mediation analysis techniques. Both my MS thesis and PhD dissertation research involved work in the field of mediation analysis. Mediation analysis is one area of causal inference that is growing increasingly popular among many fields of scientific research, and is often used to describe the underlying mechanisms through with an independent variable affects an outcome of interest. In my MS thesis work, I evaluated two new statistical tests for regression-based mediation and compared them to the two most common approaches used at the time, often referred to as *the product of coefficients* and *difference in coefficients* methods. Neither proposed method outperformed the existing methods, but we were able to demonstrate some interesting mathematical findings with respect to parameter estimate behavior with rare outcomes. In my PhD thesis, I proposed a new test for mediation based on an intersection-union test (IUT), which allows one to simultaneously test the overall effect of multiple predictor variables, similar to the concept of an overall F test on multiple regression coefficients. This new test has applicability for multiple aspects of mediation analysis, including the overall test for a single categorical variable, or the overall test of several highly correlated variables within one system (e.g. the individual components to the microbiome).

1. Wagner B, **Kroehl M**, Gan R, Mikulich-Gilbertson S, Sagel S, Riggs P, and Zerbe G (2017). A Multivariate Generalized Linear Model Approach to Mediation Analysis and Application of Confidence Ellipses. *Statistics in Biosciences*.

2. Lutz S, Thwing A, Schmeige S, **Kroehl M**, Baker C, Starling Anne, Hokanson J, Ghosh D (2017). Examining the Role of Unmeasured Confounding in Mediation Analysis with Genetic and Genomic Applications. *BMC Bioinformatics*, *In Press*.

3. Improved analytic approaches to identify risk factors in type I diabetes research. As a Research instructor for the Nutritional Etiology of Pre-Diabetic Autoimmunity, I acted as the biostatistician in a highly productive and collaborative team to identify risk factors associated with the development of Islet Autoimmunity (IA), a phase which proceeds type 1 diabetes (T1D). We collected data on the Diabetes Autoimmunity in the Young (DAISY) cohort, which has been following children at risk for T1D since 1993, including dietary behaviors using the Willet food frequency questionnaire (FFQ) and nutritional biomarkers. Because DIASY has been in existence for many years, there is a wealth of complex data which required careful statistical consideration for the analysis. A major hypothesis we had was that some risk factors had an age-related association with the development of IA; in other words, the risk factor was important during some periods of a child's life, such as during puberty, but not during others. For survival analysis, there were not existing statistical methods which allowed us to model these age-related effects using our longitudinal data (such as biomarkers collected annually) without first specifying the nature of that association (i.e. linear, quadratic, etc). I evaluated the use of restricted cubic splines using Cox regression, for use in the IVY study. This work was presented at the Joint Statistical Meeting in Montreal, Canada in 2013 and received 1st place in the Statistical Significance Poster Competition. We have since implemented this approaches in several analyses, allowing us to identify risk factors, including a SNP of growing interest among T1D researchers, which have elevated risks for development of IA during some, but not all, of childhood.

1. Frederiksen BN, **Kroehl M**, Barón A, Lamb MM, Crume TL, Sontag MK, Rewers M, Norris JM. Assessing age-related etiologic heterogeneity in the onset of islet autoimmunity. Biomed Res Int. 2015;2015:708289. doi: 10.1155/2015/708289. Epub 2015 Mar 25. PubMed PMID: 25883970; PubMed Central PMCID: PMC4389824.
2. Frederiksen BN, Steck AK, **Kroehl M**, Lamb MM, Wong R, Rewers M, Norris JM. Evidence of stage- and age-related heterogeneity of non-HLA SNPs and risk of islet autoimmunity and type 1 diabetes: the diabetes autoimmunity study in the young. Clin Dev Immunol. 2013;2013:417657. doi: 10.1155/2013/417657. Epub 2013 Dec 4. PubMed PMID: 24367383; PubMed Central PMCID: PMC3866813.
3. Frederiksen B, **Kroehl M**, Lamb MM, Seifert J, Barriga K, Eisenbarth GS, Rewers M, Norris JM. Infant exposures and development of type 1 diabetes mellitus: The Diabetes Autoimmunity Study in the Young (DAISY). JAMA Pediatr. 2013 Sep;167(9):808-15. doi: 10.1001/jamapediatrics.2013.317. PubMed PMID: 23836309; PubMed Central PMCID: PMC4038357.
4. Norris JM, **Kroehl M**, Fingerlin TE, Frederiksen BN, Seifert J, Wong R, Clare-Salzler M, Rewers M. Erythrocyte membrane docosapentaenoic acid levels are associated with islet autoimmunity: the Diabetes Autoimmunity Study in the Young. Diabetologia. 2014 Feb;57(2):295-304. doi:

10.1007/s00125-013-3106-7. Epub 2013 Nov 16. PubMed PMID: 24240437; PubMed Central PMCID: PMC3947295.

Complete List of Published Work in MyBibliography: http://www.ncbi.nlm.nih.gov/sites/myncbi/1HSOezto8j7Q_/bibliography/47461078/public/?sort=date&direction=ascending.

D. Additional Information: Research Support and/or Scholastic Performance

Ongoing Research Support

U01 AI131360 (Weinberg) 04/04/2017-03/31/2022

Tregs, Gut Microbiome and Immune Responses in HIV-Exposed and Unexposed Infants

This project will investigate the mechanism of immune regulation that controls immune responses to infections and vaccines to determine its effect on the immune response of HIV-exposed and unexposed children in their first year of life.

Role: Co-Investigator

R21 AI134129 (Weinberg) 08/08/2017-07/31/2019

The Role of Innate Immunity in the Acquisition of sterile Protection Against TB Infection

The goal of this project is to investigate immune defenses against narrative tuberculosis (TB) that may prevent the agent of TB to establish dormancy in the lungs.

Role: Co-Investigator

UL1 TR001082 (Sokol) 09/26/2013-04/30/2018

Colorado Clinical and Translational Sciences Institute

This project will re-engineer the Colorado Clinical and Translational Sciences Institute (CCTSI), funded in 2008, in order to improve efficiency, cost and effectiveness of research programs, allowing new methodologies and technologies for discovery and growth of translational success.

Role: Assistant Director of Colorado Biostatistics Consortium and Biostatistician

Completed Research Support

R33 CA183685 (Hansen) 02/05/2014 – 04/30/2017

Advanced Methods to Evaluate Extracellular Matrix and Crosslinking in the Tumor M

The goal of this project was to provide methods necessary for in-depth characterization of the extracellular matrix (ECM) involved in tumor progression to deliver previously unobtainable molecular detail regarding the microenvironment.

Role: Biostatistician

U01 HL121819 (Flores) 09/26/2013-07/31/2018

Genotypic and Functional Properties of HIV-1 NEF Clinical Isolates in PAH-HIV

This project will use existing clinical specimens and molecular tools to examine the mechanisms whereby HIV proteins influence pulmonary vascular remodeling in the pathogenesis of pulmonary hypertension associated with HIV.

Role: Biostatistician

R01 HL114887 (Stenmark) 08/15/2012-06/30/2017

NIH Axis Grant

Fibroblasts and Mononuclear Fibrogenic Cells Drive Right Ventricular Pulmonary Arterial Uncoupling in Pulmonary Arterial Pulmonary Hypertension

This project will use a combination of both human and animal studies to better understand right ventricular failure in patients with pulmonary arterial hypertension and to evaluate the potential of a novel drug therapy.

Role: Co-Investigator

U01 HL098996 (Fontenot) 09/23/2009-07/31/2015

NIH/NHLBI

Alterations in Lung Microbiome in Acute and Chronic HIV Infection

The goal of this project is to investigate relationships between alterations in the microbiome, innate and adaptive immunity, and development of chronic lung disease within HIV infected patients and health controls.

Role: Biostatistician

CHFDN 5041 (Trinkley) 10/01/2012-02/2/2015

Colorado Health Foundation Grant

Pharmacist managed Medicare outreach in primary care: lowering costs and improving health outcomes

The goal of this project is to screen and educate patients on Medicare's "extra help" programs and evaluate outcomes of such practices.

Role: Co-Investigator

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Daniel N. Frank

eRA COMMONS USER NAME (credential, e.g., agency login): DANIELFRANK

POSITION TITLE: Associate Professor of Medicine, University of Colorado

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
University of Illinois, Urbana, IL	B.S.	05/86	Honors Biology
University of Illinois, Urbana, IL	B.S.	05/86	Biochemistry
University of California, San Francisco, CA	Ph.D.	07/93	Biochemistry/Biophysics

A. Personal Statement

I will serve as a Co-PI on this proposal entitled **"Integrative Modeling of Multi-Omic Data Using a Mediation Framework"**, which will develop innovative new statistical techniques for integrating and analyzing complex 'Omics datasets (e.g., microbiome, metabolome, transcriptome) in conjunction with study outcomes. My role will be to provide expertise in the analysis of the microbiome. I currently am an Associate Professor at the University of Colorado School of Medicine, Division of Infectious Diseases. My research program explores the mechanisms by which myriad human-associated microorganisms impact our health and wellbeing. My lab has significant experience in examining human and murine microbiota in a variety of contexts, including IBD, cancer, nutrition, and HIV infection. We have conducted several large, multi-centered, international studies in both humans and rodent models that have achieved a significantly deeper understanding of the human microbiome in health and disease.

As PI or co-Investigator on several NIH and foundation-funded grants, I have developed a research group consisting of both bench scientists and bioinformaticians who collectively have the requisite skills to complete the proposed research project. I have established an excellent working relationship with Dr. Kroehl and look forward to working with her to develop these novel statistical approaches. In summary, I have a successful academic track record in both the bench-work and data analysis components of human microbiome studies. This skill set will be of critical importance to the success of the proposed project. Publications relevant to this proposal include (Key personnel in Bold):

a. Ramakrishnan VR, Hauser, LJ, Feazel, LM, Ir D., Robertson CE, **Frank DN.** Sinus microbiota varies among chronic rhinosinusitis phenotypes and predicts surgical outcome. *J. Allergy and Clin Immunol.* 2015;136(2):334-342 e331.

b. Son JS, Khair S, Pettet DW, 3rd, Ouyang N, Tian X, Zhang Y, Zhu W, Mackenzie GG, Robertson CE, Ir D, **Frank DN**, Rigas B, Li E. Altered Interactions between the Gut Microbiome and Colonic Mucosa Precede Polyposis in APCMin/+ Mice. *PLoS One.* 2015;10(6):e0127985.

c. Needell JC, Dinarello CA, Ir D, Robertson CE, Ryan SM, **Kroehl ME**, **Frank DN**, Zipris D. Implication of the intestinal microbiome as a potential surrogate marker of immune responsiveness to experimental therapies in autoimmune diabetes. *PLoS One.* 2017;12(3):e0173968.

d. Needell JC, Ir D, Robertson CE, **Kroehl ME, Frank DN**, Zipris D. Maternal treatment with short-chain fatty acids modulates the intestinal microbiota and immunity and ameliorates type 1 diabetes in the offspring. *PLoS One.* 2017;12(9):e0183786.

B. Positions and Honors

Positions and Employment

1982-86	Undergraduate, University of Illinois, Champaign-Urbana, IL
1986-93	Graduate Student, University of California, San Francisco, CA
1993-96	Postdoctoral Fellow, Indiana University, Bloomington, IN
1996-99	Postdoctoral Fellow, University of California, Berkeley, CA
1999-2010	Research Associate, University of Colorado, Boulder, CO
2010-2017	Assistant Professor, University of Colorado, Denver, CO
2013-	Affiliate, University of Colorado Center for Global Health, Denver, CO
2014-	Faculty, Graduate Program in Microbiology, University of Colorado, Denver, CO
2016-	Co-Director, Microbiome Core, GI and Liver Innate Immunity Program, University of Colorado, Denver, CO.
2017-	Associate Professor, University of Colorado, Denver, CO

Other Experience and Professional Memberships

Member, American Society for Microbiology

Member, Society for Mucosal Immunology

Member, AAAS

Member, Mucosal and Vaccine Research Program Colorado

Honors

1982-84	James Scholar, University of Illinois
1986	Phi Beta Kappa, University of Illinois
1986	Bronze Tablet, University of Illinois
1986	Summa Cum Laude, University of Illinois
1986-89	National Science Foundation Predoctoral Fellow, National Science Foundation
1993-96	American Cancer Society Postdoctoral Fellow, American Cancer Society
2013	ASPIRE Investigator Award in Adult Vaccines Research

C. Contributions to Science

Complete List of Published Work in MyNCBI:

http://www.ncbi.nlm.nih.gov/sites/myncbi/1nwB2V5UvVnQJ/bibliography/43235528/public/?sort=date&direction=descending

1. Microbiome Analysis Tools. In conducting studies of the human microbiome, my group has contributed to the development of analysis software for both 16S rRNA and meta-transcriptomic sequencing. This work has been funded through grants from the NIH Human Microbiome Project (UH2 and R21) as well as the Canadian Institutes of Health Research.

a. **Frank DN**. XplorSeq: a software environment for integrated management and phylogenetic analysis of metagenomic sequence data. BMC Bioinformatics. 2008 Oct 7;9:420. PubMed PMID: 18840282; PubMed Central PMCID: PMC2577119.

b. **Frank DN**. BARCRAWL and BARTAB: software tools for the design and implementation of barcoded primers for highly multiplexed DNA sequencing. BMC Bioinformatics. 2009 Oct 29;10:362. PubMed PMID: 19874596; PubMed Central PMCID: PMC2777893

c. Xiong X, **Frank DN**, Robertson CE, Hung SS, Markle J, et al. Generation and analysis of a mouse intestinal metatranscriptome through Illumina based RNA-sequencing. PLoS One. 2012;7(4):e36009. PubMed PMID: 22558305; PubMed Central PMCID: PMC3338770.

d. Robertson CE, Harris JK, Wagner BD, Granger D, Browne K, Tatem B, Feazel LM, Park K, Pace NR, and **Frank DN**. Explicet: graphical user interface software for metadata-driven management, analysis and

visualization of microbiome data. Bioinformatics. 2013 Dec 1;29(23):3100-1. PubMed PMID: 24021386; PubMed Central PMCID: PMC3834795.

2. Early Studies of the Human Microbiome. My studies of the human microbiome began in 1999 while a Research Associate in the lab of Prof. Norman R. Pace (University of Colorado, Boulder), who pioneered the application of culture-independent, DNA sequence-based approaches to studying microbial communities. During this time, I worked primarily on an NIH-funded study of Crohn's disease, which was one of the first very large-scale surveys (>150 subjects) of the human microbiome. The resulting paper (reference d), which demonstrated the loss of commensal *Clostridia spp.* and *Bacteroides spp.* in a subset of IBD patients, has been extensively cited. Other of my papers during this time helped to establish the basic techniques and analytic approaches that have become standards for analysis of the human microbiome.

a. **Frank DN**, Spiegelman GB, Davis W, Wagner E, Lyons E, et al. Culture-independent molecular analysis of microbial constituents of the healthy human outer ear. J Clin Microbiol. 2003 Jan;41(1):295-303. PubMed PMID: 12517864; PubMed Central PMCID: PMC149572.

b. St Amand AL, **Frank DN**, De Groote MA, Pace NR. Use of specific rRNA oligonucleotide probes for microscopic detection of *Mycobacterium avium* complex organisms in tissue. J Clin Microbiol. 2005 Apr;43(4):1505-14. PubMed PMID: 15814959; PubMed Central PMCID: PMC1081365.

c. Dalby AB, **Frank DN**, St Amand AL, Bendele AM, Pace NR. Culture-independent analysis of indomethacin-induced alterations in the rat gastrointestinal microbiota. Appl Environ Microbiol. 2006 Oct;72(10):6707-15. PubMed PMID: 17021222; PubMed Central PMCID: PMC1610281.

d. **Frank DN**, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A. 2007 Aug 21;104(34):13780-5. PubMed PMID: 17699621; PubMed Central PMCID: PMC1959459.

3. The Human Microbiome in Chronic Disease. Although disrupted gut microbiota have been demonstrated in multiple diseases, the causes and consequences of dysbiosis are only now being elucidated. The work of my group has demonstrated several mechanisms by which dysbiosis may arise and mediate disease risk: 1) Following my initial study of dysbiosis in Crohn's disease, my team of collaborators has published some of the first papers (in any disease context) linking human genetic polymorphisms with altered gut microbiota; 2) A groundbreaking study in a murine model of type 1 diabetes demonstrated that altered sex hormone expression in response to transplanted gut microbiota underlies sex-specific differences in diabetes risk; and 3) An acute trigger may alter the human gut microbiome in a manner that promotes systemic inflammation, such as occurs following HIV infection; this chronic inflammatory state results in long-term morbidity among virally-controlled HIV-infected individuals.

a. **Frank DN**, Robertson CE, Hamm CM, Kpadeh Z, Zhang T, et al. Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. Inflamm Bowel Dis. 2011 Jan;17(1):179-84. PubMed PMID: 20839241; PubMed Central PMCID: PMC3834564.

b. Markle JG, **Frank DN**, Mortin-Toth S, Robertson CE, Feazel LM, et al. Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. Science. 2013 Mar 1;339(6123):1084-8. PubMed PMID: 23328391.

c. Dillon SM, Lee EJ, Kotter CV, Austin GL, Gianella S, Siewe B, Smith DM, Landay AL, McManus MC, Robertson CE, **Frank DN,** McCarter M, Wilson CA. Gut Dendritic Cell Activation Links an Altered Colonic Microbiome to Mucosal and Systemic T Cell Activation in Untreated HIV-1 infection. *Mucosal Immunol.* 2016. 9(1):24-37.

d. Hauser LJ, Ir D, Kingdom TT, Robertson CE, **Frank DN**, Ramakrishnan VR. Investigation of bacterial repopulation after sinus surgery and perioperative antibiotics. *Int Forum Allergy Rhinol.* 2016;6(1):34-40.

4. Maternal & Infant Infectious Diseases. In addition to metagenomic analyses of the human microbiome, my lab has also characterized clinical isolates from a variety of cases of maternal and pediatric infectious diseases. Most recently, this work has entailed whole-genome sequencing of isolates, as exemplified by a molecular epidemiological study of a hospital outbreak of *C. difficile* in which my lab performed both the benchwork and data analysis.

a. Shukla SK, Vevea DN, **Frank DN**, Pace NR, Reed KD. Isolation and characterization of a black-pigmented Corynebacterium sp. from a woman with spontaneous abortion. J Clin Microbiol. 2001 Mar;39(3):1109-13. PubMed PMID: 11230435; PubMed Central PMCID: PMC87881.

b. Shukla SK, Meier PR, Mitchell PD, **Frank DN**, Reed KD. Leptotrichia amnionii sp. nov., a novel bacterium isolated from the amniotic fluid of a woman after intrauterine fetal demise. J Clin Microbiol. 2002 Sep;40(9):3346-9. PubMed PMID: 12202577; PubMed Central PMCID: PMC130742.

c. Shukla SK, Bernard KA, Harney M, **Frank DN**, Reed KD. Corynebacterium nigricans sp. nov.: proposed name for a black-pigmented Corynebacterium species recovered from the human female urogenital tract. J Clin Microbiol. 2003 Sep;41(9):4353-8. PubMed PMID: 12958268; PubMed Central PMCID: PMC193809.

d. Dominguez SR, Dolan SA, West K, Dantes RB, Epson E, eFriedman D, Littlehorn CA, Arms LE, Walton K, Servetar E, **Frank DN**, Kotter CV, Dowell E, Gould CV, Hilden JM, Todd JK. High colonization rate and prolonged shedding of *Clostridium difficile* in pediatric oncology patients. Clin Infect Dis. 2014 Aug;59(3):401-3. PubMed PMID: 24785235.

5. Phylogeny and Structure of Functional RNAs. My first exposure to the use of RNA sequences as markers of microbial ecology and cellular evolution came as an undergraduate in the laboratory of Prof. Carl Woese (Univ. of Illinois), where I isolated and sequenced several archaeal small structural RNA genes. My graduate thesis research with Prof. Christine Guthrie (University of California, San Francisco), supported by a National Science Foundation predoctoral award, used the model eukaryote *Saccharomyces cerevisiae* to probe structure/function relationships in a small nuclear RNA required for pre-mRNA processing. Following graduate school, I was awarded an American Cancer Society fellowship to pursue postdoctoral research in the laboratory of Prof. Norman Pace (University of Colorado, Boulder). There I investigated the structure and catalytic activity of the RNA subunit of ribonuclease P (RNase P), the ubiquitous and essential component of pre-tRNA processing machinery in all cellular life.

a. **Frank D**, Guthrie C. An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. Genes Dev. 1992 Nov;6(11):2112-24. PubMed PMID: 1427075.

b. **Frank DN**, Harris ME, Pace NR. Rational design of self-cleaving pre-tRNA-ribonuclease P RNA conjugates. Biochemistry. 1994 Sep 6;33(35):10800-8. PubMed PMID: 8075082.

c. **Frank DN**, Ellington AE, Pace NR. In vitro selection of RNase P RNA reveals optimized catalytic activity in a highly conserved structural domain. RNA. 1996 Dec;2(12):1179-88. PubMed PMID: 8972768; PubMed Central PMCID: PMC1369446.

d. **Frank DN**, Pace NR. In vitro selection for altered divalent metal specificity in the RNase P RNA. Proc Natl Acad Sci U S A. 1997 Dec 23;94(26):14355-60. PubMed PMID: 9405616; PubMed Central PMCID: PMC24975.

D. Research Support

Ongoing Research Support

FAMRI CIA160014 Frank (PI) 07/01/2017 – 06/30/2020

Long-term Effects of High-Intensity Secondhand Smoke Exposure on Flight Attendant Sinus Morbidity

The goal of this project is to better understand, recognize, and treat sinus disease induced by secondhand smoke (SHS) exposure a goal that is important not only to pre-ban flight attendants, but also for the 58 million Americans who continue to be exposed to SHS.

Role: PI

University of Colorado Cancer Center Lu (PI) 7/1/2016-6/30/2018

Role of oral microbiota to carcinogenesis of head and neck squamous cell carcinoma

The goal is to perform a pilot study examining oral microbiota and understand their role in head and neck cancers, using novel mouse experimental models.

Role: Co-investigator

NIH R01DK101659 Hernandez (PI) 12/01/2014 – 11/30/2019

Randomized Trial of Diet in GDM: Metabolic Consequences to Mother & Offspring

The goal of this project is to determine whether a dietary intervention relieves maternal programming of obesity and diabetes risk in infant offspring.

Role: Co-Investigator

NIH R01HD081197 Michail (PI) 7/01/2015 – 06/30/2019

Modification of Gut Microbial Profile in Children with Ulcerative Colitis

The goal of the study is to perform a randomized double-blinded placebo controlled study to examine the effect of fecal microbial transplant in disease remission and maintenance and its long-term effect on the gut microbiome and mucosal healing.

Role: Co-Investigator

NIH U01AI131360-01 Weinberg (PI) 04/04/2017 – 03/31/2022

Tregs, Gut Microbiome and Immune Responses in HIV-Exposed and Unexposed Infants

The goal of the project is to test the hypothesis that global immune dysfunction of HIV-exposed uninfected infants (HEUs) results from excessive regulatory T cells (Tregs) that develop in response to a highly inflammatory in utero environment of HIV-infected pregnant women and infant gut dysbiosis.

Role: Co-Investigator

Completed Research Support

FAMRI CIA130066 Frank (PI) 07/01/2014 – 06/30/2017

Smoke, Cellular Aging, and Chronic Rhinosinusitis

The goals of this project are to determine whether the detrimental effects of secondhand tobacco smoke on the sinuses in CRS involve the following processes: 1) Accelerated cellular aging and cell death within the tissue lining the sinuses; 2) Problems with immune function in the sinuses; and 3) Decreased amounts of protective microbes and increased amounts of pathogens (e.g. *S. aureus*) that contribute to tissue destruction and chronic inflammation.

Role: PI

University of Colorado Cancer Center Career Enhancement Grant (PI: Ramakrishnan) 9/2015-8/2017

Microbiome and inflammation in head and neck cancer carcinogenesis

In this pilot study, we will describe oral cavity microbiome alterations in patients with oral cavity and oropharyngeal cancers, and explore related mechanisms for inflammation and carcinogenesis.

Role: Co-PI

BMG-12-77-01 Gates Foundation Hambidge and Krebs (MPI) 01/1/2013-10/31/2017

Preconception Maternal Nutrition and Infant Growth

This project will be a randomized trial of a maternal nutritional supplement administered either before conception or in the first trimester. Outcomes will include changes in the development of the maternal and infant gastrointestinal microbiomes.

Role: Co-investigator

DVAMC Merit Award Bessesen (PI) 07/01/2012- 09/30/2016

Role of the nasal microbiome in methicillin-resistant *Staphylococcus aureus* carriage and infection

The goal of this project is to correlate nasal microbial communities with susceptibility to colonization and infectionby methicillin-resistant *S. aureus.* The study population will be inpatients at the Dept of Veterans Affairs Medical Center, Denver.

Role: Co-investigator

BCB 5448 Genome Canada, CIHR Parkinson (PI) 01/01/2014 – 06/01/2016

Leveraging Meta-Transcriptomics for Functional Interrogation Of Microbiomes

We are developing an innovative software platform that combines accurate transcript annotation with systems-level functional interrogation of meta-transcriptomic datasets.

Role: Co-investigator

NIH R21AA022387 Ju (PI) 9/20/2013 – 08/31/2016

Effect of Lactoferrin on Alcohol-Induced Dysbiosis and Gut Barrier Dysfunction

The goal of this project is to determine whether oral administration of lactoferrin ameliorates alcohol-associated alterations in murine gut microbiome and epithelial barrier function.

Role: Co-Investigator

Pfizer Inc. WI182737 Frank (PI) 11/01/2013 – 06/30/2015

Impact of the human nasopharyngeal microbiome on *Streptococcus pneumoniae* carriage in immunocompromised individuals

The goals of this study are to determine whether HIV-infected individuals are colonized by different pneumococcal serotypes as well as competing commensal microorganisms, compared with non-infected subjects.

Role: PI

University of Colorado
School of Medicine

Department of Otolaryngology

Mail Stop B205
12631 East 17th Avenue, Room 3001
Aurora, CO 80045
Office: 303-724-1950
Fax: 303-724-1961

January 11, 2018

Dear Miranda:

We would like to express our enthusiasm for the proposal, "Integrative Modeling of Multi-Omic Data Using a Mediation Framework" that you and Dan are submitting to the CCTSI NMD Pilot Program. This work will fill an important existing gap in available statistical tools and allow for better integration and statistical modeling to evaluate microbiome – metabolite – disease pathways.

These tools will be invaluable in our research on head and neck squamous cell carcinomas (HNSCC) in which we know little about the role of the microbiome in disease development, and less about the metabolic pathways through which microbiota may act. Over the past few years, we have established a cross-sectional cohort of over 150 HNSCC patients and health controls and have been collecting samples, including saliva, for microbiome and metabolite profiling. We have completed 16S rRNA sequencing on these subjects though Dan's lab, and have generated metabolite data on 20 samples to be used as pilot data for some upcoming grant proposals. Within the first half of this year, we will be subjecting additional samples to untargeted metabolomic analysis, and anticipate there will be a cohort of nearly 150 subjects (accounting for missing samples, poor quality, etc.) for which we have microbiome, metabolome, and patient outcome data. This will be an extremely valuable set of data to evaluate under this new integrative framework.

The findings from your analyses will uncover new relationships between the microbiome and metabolome and identify specific bacteria and metabolites which warrant further investigation. These findings and resulting joint publications will increase the competitiveness of future grant proposals, including upcoming R01 and SPORE proposals we have planned. Furthermore, the hypothesis that the microbiome affects disease outcomes through the metabolome is not one that is limited to HNSCC. In addition to the HNSCC cohort, we also have microbiome and immune data from an existing study of chronic rhinosinusitis (CRS), and anticipate that the analytic framework you develop will be of great value in integrating and evaluating these datasets moving forward. We also anticipate that many other colleagues on campus will find this framework to be quite useful, as the microbiome is now seemingly investigate in nearly every disease process. *In fact, the single most important external review commentary we have received, both from the SPORE External Advisory Board and at a pan-institute Microbiome Symposium at NIH, was the need for integration of the microbiome with functional 'omics data sets.*

We have enjoyed working with you on our current HNSCC and CRS endeavors, and as you know, plan to continue this collaboration on future projects, including the aforementioned grant proposals. We are excited to support this current proposal by providing metabolome samples and look forward to working with you.

Sincerely,

Vijay Ramakrishnan, MD
Associate Professor
Department of Otolaryngology
University of Colorado Anschutz Medical Campus

Shi-Long Lu, MD, PhD
Associate Professor
Department of Otolaryngology
University of Colorado Anschutz Medical Campus