DEBUGGING LUNG DISEASES: APPLYING MATHEMATICAL TECHNIQUES, INVOLVING MODELLING, DATA INTEGRATION AND MACHINE LEARNING FOR PRECISION MEDICINE

QUALIFYING EXAMINATION REPORT

SUBMITTED BY: JAYANTH KUMAR NARAYANA
MATRICULATION NUMBER : G1902804D

Lee Kong Chian School of Medicine, Singapore

Contents

List of Figures

List of Tables

1   Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc,

Introduction

The term microbiome is used to refer to the collection of genes within a community of microbes (including bacteria, fungi, virus, protists and bacteriophages). In the last few years, microbiome research has helped us gained new insights into how microbes shape our human biology and have brought paradigm-shifting implications for translational research and clinical care. The human microbiota is crucial for our body to maintain its homeostasis. Disruption of this can lead to diseases such as obesity, inflammatory bowel disease, malnutrition, Parkinson's, Autism, Asthma, dental caries, bacterial vaginosis, and depression [3]. Currently, microbiome researchers use culture-independent techniques that involve DNA sequencing to derive the microbiome. Broadly, the community taxonomy/microbiome can be identified using two approaches (see Figure1) 1) Targeted and 2) Metagenomic. Targeted sequencing approach uses the PCR amplified, target gene markers (16S rRNA in case of bacteria or ITS in case of Fungi) derived from the samples to reference it against gene-marker databases (Silva, Green Genes, etc.). In contrast, the metagenomic sequencing approach directly sequences the whole community DNA and compares it to reference genomes [5].
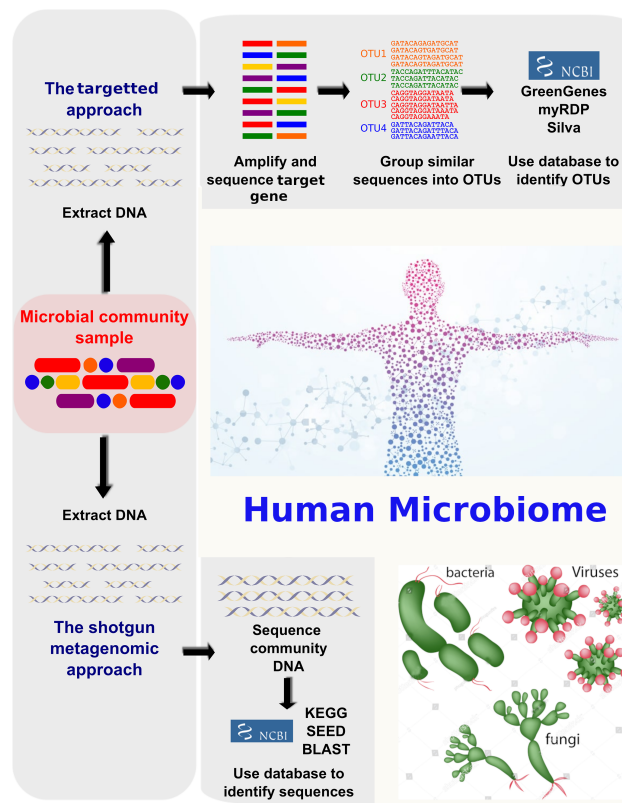
Figure 1: A figure illustrating the different sequencing approaches used to derive the human microbiome, consisting of interacting bacteria, fungi and viruses. Adapted from: [5]

Present microbiome studies focus on a single profile of the human microbiome in isolation, even though bacteria, fungi and viruses coexist and interact in the body as a community. Thus, it is essential to look at these biological components together in an integrated fashion to understand more holistically the true underlying in vivo state. However, one of the primary reasons for the lack of multi-biomic research is the lack of methods to merge microbiome datasets and integrative analysis. Consequently, I tried addressing some of these challenges in my master's thesis, using microbiome datasets derived from bronchiectasis patients as an example. Bronchiectasis, is a chronic inflammatory respiratory disease associated with progressive, irreversible dilatation of the airway. It is crucial to study bronchiectasis because in most cases it is known to be idiopathic(unknown cause) [1] and it is a significant contributor to lung diseases globally with a substantial four-fold higher predominance in Asian populations [7].
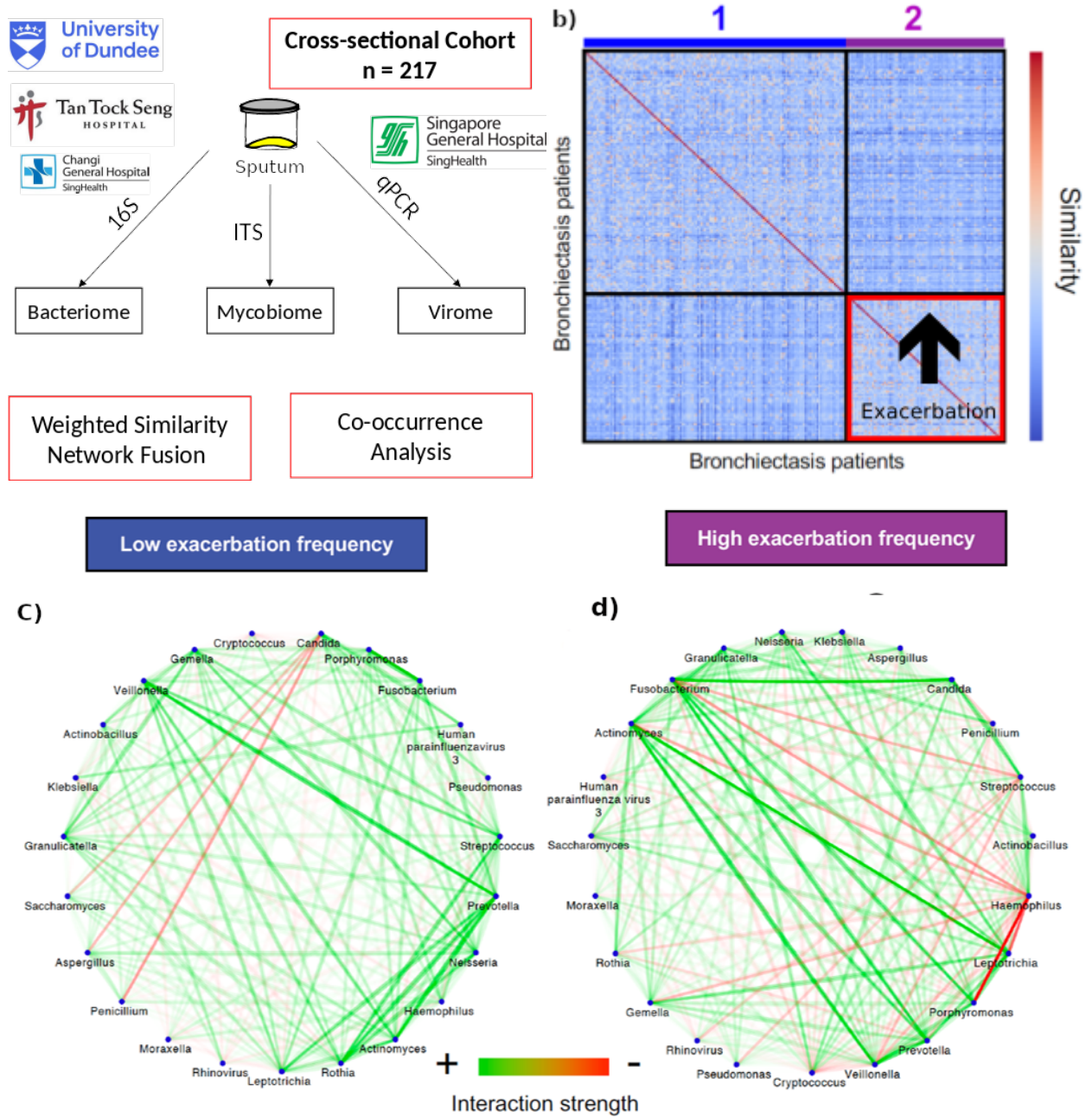
Figure 2: (a) A schematic representing, overview of analysis performed on the CAMEB cohort (n=217). Methodologies: Weighted SNF and Co-occurrence analysis were used for microbiome integration and intreactome construction. (b) A patient similarity matrix with each cell representing the integrated similarity between patients. Two clusters of low (black) and high (red) risk patients identified by wSNF are highlighted by boxes. Visualization of the interactome of low (c) and high (d) risk clusters. Interactions between microbes are classified as negative if the sign of the edge weights between them is negative (coloured red) with positive interactions indicated by green colouration. The strength of the interaction is indicated by the colour depth

Previously in my master's thesis, we developed weighted similarity network fusion (wSNF) to allow weightage of input datasets during integration, otherwise unaccounted by conventional SNF [8]. Ensemble-based co-occurrence analysis strategy developed by Faust et al. [2] was extended to allow weightage of individual methods in the ensemble along with other modification to better infer microbial association networks. Microbiome and Mycobiome derived using targeted amplicon sequencing of the 16S and ITS regions from the sputum samples of the CAMEB cohort [4]; virome from qPCR on an extensive panel of 17 respiratory viruses, were used as the example dataset to integrate the microbiomes (Figure2a). Multi-biome (Microbiome, Mycobiome and Virome) integration by wSNF identifies a high-risk exacerbation cluster with increased precision (Figure2b). Co-occurrence network analysis of this high-risk cluster revealed an elevated

antagonistic interactome with reduced alpha-diversity (Figure2c) [6].

Having developed the wSNF and shown its increased precision to identify exacerbators (clinical outcomes); here in my PhD thesis, I attempt to extend my results further. We aim to develop a web tool to enable users to integrate their microbiome datasets and to illustrate its advantages using publicly available microbiome datasets. The tool would motivate clinicians and microbiome researchers to explore multi-biome strategies for their problem and aid them in integrating their datasets. Secondly, we aim to study exacerbation events, antimicrobial perturbations and "Time to next exacerbation" using the developed "Interactome" framework. Thirdly, we aim to validate the "high-risk" exacerbation cluster of Bronchiectasis patients and its "interactome" as derived in my previous work [6] using an alternate sequencing approach: metagenomics. Further, we also pick an interaction from the interactome of the high and low-risk clusters and validate it experimentally.
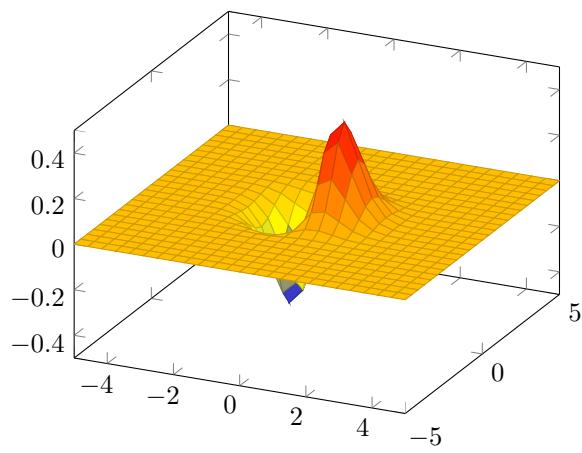
2 Methods

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc, quis gravida magna mi a libero. Fusce vulputate eleifend sapien. Vestibulum purus quam, scelerisque ut, mollis sed, nonummy id, metus. Nullam accumsan lorem in dui. Cras ultricies mi eu turpis hendrerit fringilla. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; In ac dui quis mi consectetuer lacinia. Nam pretium turpis et arcu. Duis arcu tortor, suscipit eget, imperdiet nec, imperdiet iaculis, ipsum. Sed aliquam ultrices mauris. Integer ante arcu, accumsan a, consectetuer eget, posuere ut, mauris. Praesent adipiscing. Phasellus ullamcorper ipsum rutrum nunc. Nunc nonummy metus. Vestibulum volutpat pretium libero. Cras id dui. Aenean ut

Test reference here. [**?**]
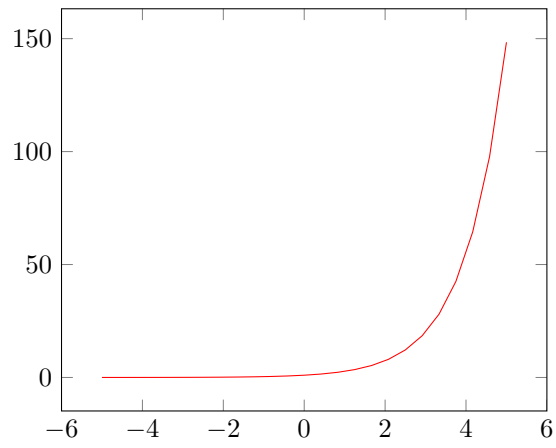
Figure 3: Plot of Exponential Function.

Figure 4: Plot of 3-D Function.

3   Results

3.1   Math Cheat Sheet

Figure 5: Two functions in one plot.

Figure 3, Figure 4, and Figure 5 are generated from function in LaTeX.

3.2   Table Cheat Sheet

	Apple	Oranges	Strawberries
A	1	2	3
B	1	2	3
C	1	2	3
D	1	2	3

Table 1: Caption

4   Discussion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc, quis gravida magna mi a libero. Fusce vulputate eleifend sapien. Vestibulum purus quam, scelerisque ut, mollis sed, nonummy id, metus. Nullam accumsan lorem in dui. Cras ultricies mi eu turpis hendrerit fringilla. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; In ac dui quis mi consectetuer lacinia. Nam pretium turpis et arcu. Duis arcu tortor, suscipit eget, imperdiet nec, imperdiet iaculis, ipsum. Sed aliquam ultrices mauris. Integer ante arcu, accumsan a, consectetuer eget, posuere ut, mauris. Praesent adipiscing. Phasellus ullamcorper ipsum rutrum nunc. Nunc nonummy metus. Vestibulum volutpat pretium libero. Cras id dui. Aenean ut
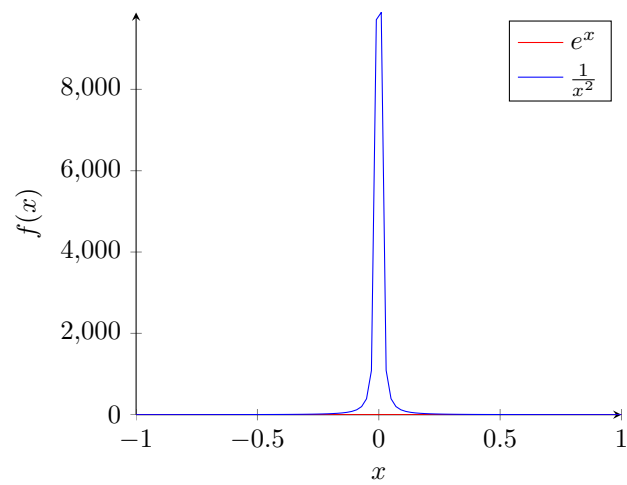
5 Future works

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc, quis gravida magna mi a libero. Fusce vulputate eleifend sapien. Vestibulum purus quam, scelerisque ut, mollis sed, nonummy id, metus. Nullam accumsan lorem in dui. Cras ultricies mi eu turpis hendrerit fringilla. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; In ac dui quis mi consectetuer lacinia. Nam pretium turpis et arcu. Duis arcu tortor, suscipit eget, imperdiet nec, imperdiet iaculis, ipsum. Sed aliquam ultrices mauris. Integer ante arcu, accumsan a, consectetuer eget, posuere ut, mauris. Praesent adipiscing. Phasellus ullamcorper ipsum rutrum nunc. Nunc nonummy metus. Vestibulum volutpat pretium libero. Cras id dui. Aenean ut [**?**]

References

[1] CHALMERS, J. D., AND CHOTIRMALL, S. H. Bronchiectasis: new therapies and new perspectives. *Lancet Respir Med 6*, 9 (Sep 2018), 715–726.

[2] FAUST, K., SATHIRAPONGSASUTI, J. F., IZARD, J., SEGATA, N., GEVERS, D., RAES, J., AND HUTTENHOWER, C. Microbial co-occurrence relationships in the human microbiome. *PLoS computational biology 8*, 7 (2012), e1002606.

[3] KNIGHT, R., CALLEWAERT, C., MAROTZ, C., HYDE, E. R., DEBELIUS, J. W., MCDONALD, D., AND SOGIN, M. L. The microbiome and human biology. *Annual Review of Genomics and Human Genetics 18*, 1 (aug 2017), 65–86.

[4] MAC AOGÁIN, M., CHANDRASEKARAN, R., LIM YICK HOU, A., TECK BOON, L., LIANG TAN, G., HASSAN, T., THUN HOW, O., HUI QI NG, A., BERTRAND, D., YU KOH, J., LEI PANG, S., YANG LEE, Z., WEI GWEE, X., MARTINUS, C., YIE SIO, Y., ANUSHA MATTA, S., TIM CHEW, F., KEIR, H. R., CONNOLLY, J. E., ARPUTHAN ABISHEGANADEN, J., SIYUE KOH, M., NAGARAJAN, N., CHALMERS, J. D., AND CHOTIRMALL, S. H. Immunological corollary of the pulmonary mycobiome in bronchiectasis: The cameb study. *European Respiratory Journal* (2018).

[5] MORGAN, X. C., AND HUTTENHOWER, C. Human microbiome analysis. *PLoS Comput Biol 8*, 12 (2012), e1002808.

[6] NARAYANA, J. K. Investigating the respiratory microbiome in bronchiectasis through "integrative microbiomics". Master's thesis, Indian Institute of Science Education and Research, Pune, 2019.

[7] SEITZ, A. E., OLIVIER, K. N., ADJEMIAN, J., HOLLAND, S. M., AND PREVOTS, D. R. Trends in bronchiectasis among medicare beneficiaries in the united states, 2000 to 2007. *CHEST 142*, 2 (Aug 2012), 432–439.

[8] WANG, B., MEZLINI, A. M., DEMIR, F., FIUME, M., TU, Z., BRUDNO, M., HAIBE-KAINS, B., AND GOLDENBERG, A. Similarity network fusion for aggregating data types on a genomic scale. *Nature Methods 11* (Jan 2014), 333 EP –. Article.