Creating an Interactive Dashboard for Tuberculosis and its Top Mutations

Sudheshna Bodapati & Monét Norales

1. Introduction

Tuberculosis (TB) is an infectious bacterial disease caused by a bacteria called mycobacterium tuberculosis. It is commonly known for having the symptom of the infected person coughing up blood. However, it can cause disease in various areas of the respiratory, immune, and nervous systems. TB is still one of the highly threatening health risks in third world countries killing over about a million people each year. TB is treated with early-stage antibiotics such as rifampicin, streptomycin, fluoroquinolones, etc. The researchers are facing significant challenges in controlling TB due to the mutated and drug-resistant strains, especially the strains that have developed (MDR) Multi-drug Resistant properties. The first-line antibiotic therapies are considerably less efficacious coupled with a significant gap between those in need of the second-line therapy and those who receive it, reminding us of the need for new approaches immediately to tackle this MDR problem. We sought to summarize and visualize the available evidence on genotype changes for the top 15 mutations with MDR across the world.

2. One Sentence Description

Based on information found from various government databases and other research papers, the goal is to generate a database schema to visualize the top 10 mutations at different genes, their base pair locations, type of drug resistance, and their epidemiological impact.

3. Project Type

Dashboard / Database / Webpage

4. Audience

The audience will be the biology and bioinformatics early researchers or maybe academia within those fields. As the community searches for a refined drug resistance mutational database to understand its impact on both genomic level and population level they should be able to find this dashboard/website to be a decent reference. This has the potential to be a great head start to further conducting a deep analysis of a genomic, proteomic, or transcriptomic analysis. This could also be used for a more simple population health analysis and overview. This in turn is to

encourage the research community to look more into TB and gain a better understanding of the main mutations that occur.

If the needs are unmet, there it will be harder to understand the datasets given by WHO and the CDC. With this, it could decrease the interest in further research if the data is difficult to manipulate and investigate. There is also potential for increased difficulty in noticing and addressing trends or comparisons between the various mutations and their influences.

4.1. Approach Details

When approaching this project, first the data will be retrieved from the WHO and other published journals and databases that give us the mutations correlated with the drug resistance and analyze them.
We are interested in understanding the relationship between the mutations spread within the genomic level and their spread across different nations.

4.2. Evidence for Success

There is a vast amount of research in the field of TB and there are several websites that give us a reference as to how to build a new one. This project should further investigate and try to integrate only the most useful data considering the short frame of time.

Two major epidemiological databases WHO and CDC can help us identify the TB statistics within different socio-economic factors. Three major databases, Tuberculosis Drug Resistance Mutation Database (TBDReaMDB), Mutation Bioinformatics Identification (MUBII-TB-DB), Tuberculosis Drug Resistance Database (TDRB) are currently available for mutations associated with drug resistance in TB.

5. Best-Case Impact Statement

The best scenario is that the dashboard is functional and the displays accurately represent that data. Also that there is access and accurate information given about the top 10 mutations which are to be represented.

6. Major Milestones

The main milestones to accomplish for this project are to first find the relevant datasets and clean the dataset. Then the manipulation possibly with SQL databases to host the data in the background and connect it to the GitHub repository. Also, the database would need to be connected to the javascript. Finally, the results need to be displayed in a manner that is easy for the general audience to understand and has filtering functions and other attributes.

7. Obstacles

7.1. Major Obstacles

Major Challenges when making this dashboard include: working with a d3 library on such a big scale and possibly contacting a TB expert/researcher to ensure that the data is well interpreted and conveyed.

7.2. Minor Obstacles

Some of the minor challenges that may be encountered include: creating various interaction points within the main dashboard, adding hover effects to emphasize specific information, and figuring out the best visualization method for depicting the different types of information.

8. Resources Needed

To complete the desired goal, there might be a need for help in the linking of the various databases or adding in more complex interaction functions.

9. 5 Related Publications

- The datasets in [5] will be a baseline for some (if not all) of the created visualizations.
- The information in [1] and [3] provide insight into the trends and infection rates for the past years.
- More about some the mutation rates and emergence in [4] which can be helpful in the depictions of the mutations and the finer details about them.
- From [7] another dataset is available, also there is more about the analysis of sequences that can be used when describing and understanding the mutations.
- Paper [6] gives an important insight as to the prediction and ranking of mutations which can be used when determining the ranking of the top 10 mutations if there is consideration beyond which is more prevalent.

10. Define Success

A minimum successful outcome for this project should have a well-organized dashboard that has multiple diagrams that are interactive and linked to the datasets available on the WHO website.

11. References

[1] Centers for Disease Control and Prevention. (2021, October 12). *Trends in Tuberculosis, 2020*. Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/tb/publications/factsheets/statistics/tbtrends.htm

[2] World Health Organization. (n.d.). *Tuberculosis (TB)*. World Health Organization. Retrieved from https://www.who.int/news-room/fact-sheets/detail/tuberculosis

[3] Ford, C., Shah, R., Maeda, M. *et al.* *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of

drug-resistant tuberculosis. *Nat Genet* 45, 784–790 (2013). https://doi.org/10.1038/ng.2656

[4] Countries Reporting to WHO. (n.d.). *Global Tuberculosis Report*. World Health Organization. Retrieved from https://www.who.int/teams/global-tuberculosis-programme/data

[5] Kouchaki, S., Yang, Y., Lachapelle, A., Walker, T. M., Walker, A. S., Consortium, C. R. P. T. I. C., Peto, T. E. A., Crook, D. W., & Clifton, D. A. (1AD, January 1). *Multi-label random forest model for tuberculosis drug resistance classification and mutation ranking*. Frontiers. Retrieved from, https://www.frontiersin.org/articles/10.3389/fmicb.2020.00667/full

[6] Kohl TA, Utpatel C, Schleusener V, De Filippo MR, Beckert P, Cirillo DM, Niemann S. 2018. MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex isolates. *PeerJ* 6:e5895 https://doi.org/10.7717/peerj.5895

[7] Koch, A., Cox, H., & Mizrahi, V. (2018, June 6). *Drug-resistant tuberculosis: Challenges and opportunities for diagnosis and treatment*. Current Opinion in Pharmacology. Retrieved from, https://www.sciencedirect.com/science/article/pii/S1471489217301571

[8] Ghosh, A., N., S., & Saha, S. (2020, June 2). *Survey of drug resistance associated gene mutations in mycobacterium tuberculosis, Eskape and other bacterial species*. Nature News. Retrieved from, https://www.nature.com/articles/s41598-020-65766-8

[9] Ismail, N., Omar, S. V., Ismail, N. A., & Peters, R. (2018). Collated data of mutation frequencies and associated genetic variants of bedaquiline, clofazimine and linezolid resistance in *Mycobacterium tuberculosis*. *Data in brief*, *20*, 1975–1983. https://doi.org/10.1016/j.dib.2018.09.057

[10] Colangeli, R., Gupta, A., Vinhas, S.A. *et al. Mycobacterium tuberculosis* progresses through two phases of latent infection in humans. *Nat Commun* 11, 4870 (2020). https://doi.org/10.1038/s41467-020-18699-9

[11] Kadura, S., King, N., Nakhoul, M., Zhu, H., Theron, G., Köser, C. U., & Farhat, M. (2020, May 3). *Systematic review of mutations associated with resistance to the new and repurposed mycobacterium tuberculosis drugs bedaquiline, clofazimine, linezolid, Delamanid and Pretomanid*. OUP Academic. Retrieved from, https://academic.oup.com/jac/article/75/8/2031/5828363