Multiple and Extensively Drug Resistant Tuberculosis Data Exploration Portal (MXDR-TB DEPot) Examples: Cohort Creation, Visualization, Analysis

Andrei Gabrielian, Eric Engle October 2016

Table of Contents

Introduction

Example 1: Finding patients with contradictory Hain versus Lab Cultures.
Example 2: Finding features essential for survival for patients with XDR TB.
Example 3: Characterize a user uploaded X-ray image by finding the similar images in the database.

Table of Figures

Figure 1 Bactec sensitive ad Hain test resistant
Figure 2 Bactec resistant and Hain test sensitive
Figure 3 Summary View of XDR cured versus XDR died or failed
Figure 4 Detailed statistics for Lung Capacity Decrease
Figure 5 Diseased lung image (with cavities) similarity search result
Figure 6 Clinical description: resistance and outcomes for the similar images cohort
Figure 7 Summary view of similar images cohort versus all other cases
Figure 8 Detailed Statistics for Post TB residuals
Introduction
The multiple and extensively drug resistant data exploration portal (MXDR-TB DEPot) enables users to organize information from tuberculosis resources, including the country-specific TB portals, to help in multi-factor analysis of rare, unusual, atypical, most dangerous MXDR-TB cases. This document provides sample guided analyses to aid users and provide an initial introduction to the mechanics of using the DEPot system to create cohorts, visualize and analyze data.

Cohort creation is DEPot functionality that enables the selection of a subset of patients from the TB Portals data collection using data filters connected with logical operators if needed. The DEPot can for many purposes.

- To discover what is unusual in terms of outcomes and resistance
- Browse for similar patient cases and read the patients’ histories – as if consulting with a referenced patient case folder
- View the most interesting features for the cohort by clicking on the tabs for Clinical features, CT Scans, or Genomic features
- To compare any two cohorts out of the collections of patient cases
- To visualize the data and conduct statistical analysis to test hypotheses between two cohorts
Example 1: Finding patients with contradictory Hain versus Lab Cultures.

Aim: Provide user-friendly interface allowing to combine rich metadata descriptors to find use cases that require additional analysis or constitute interesting cohorts for sequencing.

Description: DEPOT can be used to find patient(s) that exhibited unusual symptoms, treatment results, and contradictory results of resistance testing or lab tests.

We want to find patients that have information about both Bactec and Hain resistance diagnostics. Bactec is based on growing the culture in the presence of the drug and Hain is based on SNP presences (PCR based test). If the results of these two tests do not coincide it probably means that the actual resistance was not correctly predicted by Hain.

---

Figure 1 Bactec sensitive ad Hain test resistant
Figure 2 Bactec resistant and Hain test sensitive

<table>
<thead>
<tr>
<th>Queries utilized</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bactec_test: s AND hain_test: r</td>
<td>Bactec test for the sample indicates a drug sensitive bacterial infection AND Hain test for the same sample indicates a drug resistant bacterial infection</td>
</tr>
<tr>
<td>bactec_test: r AND hain_test: s</td>
<td>Bactec test for the sample indicates a drug resistant bacterial infection AND Hain test for the same sample indicates a drug sensitive bacterial infection</td>
</tr>
</tbody>
</table>

The aforementioned queries take advantage of logical operator ‘AND’ to seek patients with contradictory results of Bactec test culture and Hain test. There are 14 cases when Hain test predicted resistance and Bactec test did not confirm it. There is only 16 case when Hain test did not confirm a positive result of the Bactec test. Overall these 30 cases might constitute a good cohort for further full genome sequencing to find out whether known and novel SNPs might explain such contradiction. In order to improve diagnostics tools it is useful to find contradictory cases for further study of genomic variability.

**Example 2: Finding features essential for survival for patients with XDR TB.**

**Aim:** Contribute to best practices in patient care by analyzing clinical metadata correlating with treatment outcomes.

**Description:** DEPOT can be used to find patients for whom the treatment was unsuccessful. There could be multiple factors contributing to negative outcome. Identifying the most important and manageable parameters might help physicians’ to prepare effective treatment plans.
We want to find all patients with extremely drug resistant Tuberculosis (XDR-TB) and then divide them into two groups. The first group consisted of patients that were successfully treated. Patients in the second group either failed the treatment or died.

<table>
<thead>
<tr>
<th>Queries utilized</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>type_of_resistance: xdr AND outcome: cured</td>
<td>XDR (extremely drug resistant TB) patients that were cured</td>
</tr>
<tr>
<td>type_of_resistance: xdr AND (outcome: failed OR outcome: died)</td>
<td>XDR (extremely drug resistant TB) patients that either failed the treatment or died</td>
</tr>
</tbody>
</table>

The DEPOT then will identify the most statistically significant differences between these two groups of patients.

*Figure 3 Summary View of XDR + cured versus XDR + (died or failed)*
The DEPOT analytics illustrate the differences between Cohort A: XDR and Cured (25 cases) and Cohort B: XDR and (Failed or Died) (20 cases). Note that age of onset and gender were found as important descriptors to distinguish between the two cohorts. However, in terms of monitoring and disease prognosis, the decrease in lung capacity is the only feature that might be mitigated with a treatment plan. It appears that, for survival of patients with XDR-TB, decreasing lung capacity may be predictive of outcome.

This finding is in agreement with the published results of Chen et al, “PET/CT imaging correlates with treatment outcome in patients with multidrug-resistant tuberculosis”. The article’s abstract notes the following.

- CT scans at 6 months (but not 2 months) assessed by radiologist readers were predictive of outcomes, and changes in computed abnormal volumes were predictive of drug response at both time points.
- In this cohort, some radiologic markers were more sensitive than conventional sputum microbiology in distinguishing successful from unsuccessful treatment.
Example 3: Characterize a user uploaded X-ray image by finding the similar images in the database.

Aim: Use the automatically generated image similarity as a way to access the content of TB Portals database and start statistical analysis of clinical features.

Description: DEPOT enables users to bring in an X-ray image of lungs to compare with existing annotated lung images and consequently form cohort(s) out of the computed most similar images. The comparison is done automatically based on algorithm developed by Dr. Vissaly Kovalev’s lab and adapted by Octavio Juarez-Espinosa.

The user can browse the most similar images and read the patients’ histories – as if they were consulting with a referenced patient case folder. The user might browse all other clinical features for the cohort (by clicking on the tabs for Clinical features, CT Scans, or Genomic features). Finally, the user may retain the cohort of the original image and most similar images and conduct comparative analysis against other cohorts.

Using the CT images tab the user can access detailed descriptions of lungs provided by qualified radiologists. For example, for an uploaded image (found on the Internet with the keyword “cavity mass”), most similar images from TB Portals database include patients with massive lung damage (more than two segments). Finding the successfully treated patients in the cohort might suggest possible treatment strategy. Conversely, finding unsuccessfully treated patients in the cohort might help in identifying treatment strategies to avoid.
Figure 5: Diseased lung image (with cavities) automated similarity search result – 9 most similar images were picked from the database and presented for further analysis.

<table>
<thead>
<tr>
<th>COMMON FILTERS</th>
<th>GENOMICS</th>
<th>CT SCANS</th>
<th>X-RAYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>2.4%</td>
<td>97.6%</td>
<td>367</td>
</tr>
<tr>
<td>Comorbidity</td>
<td>12.1%</td>
<td>87.9%</td>
<td></td>
</tr>
<tr>
<td>Diagnosis Code</td>
<td>10.3%</td>
<td>89.7%</td>
<td></td>
</tr>
<tr>
<td>Type of Resistance</td>
<td>10.3%</td>
<td>89.7%</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>5.5%</td>
<td>94.5%</td>
<td></td>
</tr>
<tr>
<td>Case Definition</td>
<td>6.0%</td>
<td>94.0%</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>0.1%</td>
<td>99.9%</td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td>0.0%</td>
<td>100.0%</td>
<td>184</td>
</tr>
</tbody>
</table>

Figure 6: Clinical description: resistance and outcomes for the similar images cohort.
Figure 7  Summary view of similar images cohort versus all other cases

Figure 8 Detailed Statistics for Post TB residuals