Stop TB Partnership

New Diagnostics Working Group

Critical Path to TB Drug Regimens

Coordination of TB diagnostics research: Enabling standards and sharing of data on the molecular basis of drug resistance

A workshop jointly organized by the New Diagnostics Working Group and CPTR

Senate House, London, UK
3 - 4 February 2014

WORKGROUP 1

DATA SETS AND PILOT PROJECTS

Chair: Matteo Zignol, WHO

Rapporteur: David Alland, Rutgers-NJMS

Objectives

- Establish a preliminary list of existing databases and groups that own or manage them *(a paper from LSHTM in this regard will be circulated before the meeting, along with others as available)*
- Identify the datasets that are available and what is the content of those data /
- What has been sequenced (i.e. processed sputum, strains, etc.) and what was defined as resistant (i.e. the phenotype, the genotype, the clinical outcome)
- Determine what science questions linked to detection of drug resistance are left unanswered (which markers, mutations predictive of DR, etc.) and for which reasons (operational, structural, scientific)
- Agree on what data need to be collected to answer those questions and fill the gaps

Data sets to use

- Single read archive available at EBI: 1200 from the Midlands, 300 Oxforshire, 2500 Malawi.
- Singapore genome center: 3000 from Asia.
  DST available for much of this for most first line.
- Population-based Hamburg study: 10 years of TB cases.
  Also 100 MDR and 100 DS strains looking for mutations associated with DR.
- TB-CDRC/Broad: 200 isolates MDR/XDR.
  TREK MIC and agar proportion.
- WHO PZA surveillance project: 5000 isolates from 5 countries.
  Full genome in South Africa only.
- Roetzer et. al, PLoS medicine: 96 samples, longitudinal outbreak in Northern Germany.
- Megan Murray's group: 500 strains with another 1500 strains.
  Most drug resistant and hopefully with MIC data.
- Sanger: 1000 strains strains from Sweden, Peru, South Africa.
- Sanger: Part of 100,000 genome project. Several thousand TB strains.
- Must also be done in China and India. Need to get this information.

What questions need to be answered for diagnostics? (1/2)

- Link genotype to phenotype.
  This is still an important area to study. Not necessary to specify what type of DST but list what system used and what breakpoints.
  Essential to include basic demographic data to test data sets for inclusiveness (representation across geography and time).
- Minimal meta data.
  Unique identifier, country of origin, DST (to at least some drugs), information about DST (breakpoint, media), specimen type, date of collection, sequencing platform, submitter ID.
- Large data sets can be used to generate hypothesis to get to the next level. This could be done with relatively simple data fields.
  DST, relapse (Y or N), number of relapses, time since TB treatment?
- Focus to get data available at the time the sample is collected is the surest way to get a complete dataset.
- We should also seek to get outcome data.
  This would be enormously valuable and worth the massive investment.

<u>What questions need to be answered for diagnostics?</u> (2/2)

- Link genotype to clinical outcome.
  This type of database may need to be better curated and have limited access.
- Link with PK data if available.
  Gender, age, ethnicity data would also be important to interpret PK results.
- How do you get clinical outcome into management of XDR TB cases? Model of HIV? But may be more difficult because of more transmitted resistance in TB.
- Pull together data from MDR/XDR studies such as Remox.
- Can test existing datasets to see if they meet minimal standards.
- Database should have some pre analysis:
  R vs S for each drug (if MIC supplied), clade.
- Do we need a second data base that includes something like only SNPs of interest (model is HIV resistance database)?
  Alternative is to have database with whole genomes with pipelines that feed into a relational database that interfaces with a user-friendly website.
  But in terms of resources, need to realize that these two goals are separate programing problems.