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

Data sharing for TB diagnostics: Needs and gaps

Marco Schito, HJF-DAIDS (NIH)
Rapid TB DST Consortium

Mission:

To accelerate the development of a **clinically** useful, **WHO-qualified**, regulatory approvable IVD assay for rapid TB-DST

• Such an assay is needed to support the
  - Enrolment of volunteers in clinical trials (inclusion/exclusion)
  - Commercial assay development
  - Optimize role out of new regimens (PaMZ, REMox...)

• Push for nucleic acid diagnostics to meet short TAT

• Need for a global TB drug resistance database to house genomic data for new and existing drugs
  - Who is going to use the database and why?
  - What questions do we want to address (today, 5, 10 years)?
  - Challenges for new, existing and repurposed drugs and regimens?
1. **Research and development**
   - Level of evidence that a particular mutation correlates with resistance
   - Define sensitivity (modeling clinical impact vs. cost)

2. **Ongoing global surveillance**
   - Resistance data beyond HRES
   - Inform public health and clinical trials (power calc.)
   - Identify market needs and guide diagnostic algorithms

3. **Biomarkers**
   - Identification of a signature(s) associated with disease progression or response to therapy

4. **Clinical management**
   - Guide treatment decisions based on sequence information
   - Which mutations are clinically relevant?

5. **Regulatory**
   - Compliant with health authority requirements for IVD claims
Database structure for existing drugs

Koser et al. 2012. Plos Path. 8(8) e1002824
Legal issues regarding new drugs

• Pharmaceutical companies need to:
  - Include all information to authorities prior to regulatory submissions
  - Protect IP issues regarding future diagnostic applications
  - Restrict:
    a. Drug compounds for diagnostic evaluation
    b. Clinically relevant data important for developing a diagnostic DST

• CPTR effort to develop:
  - A framework whereby all stakeholders can collaborate, address concerns from drug companies and comply with regulatory policies
  - A “protected” database for a select group of participants who have scientific, technical or clinical expertise to help advance diagnostic assays for new drugs
  - Basis of the legal agreement
Proposed database for new/existing drugs

New Drugs
- Legal agreement

Key participants
- Clinical trial groups
- Pharmaceutical Industry
- Diagnostic developers
- Scientists/clinicians
- Governments
- NGO

Clinical Trial data repository
- New TB drugs
- New regimens
- Pheno/geno
- Clinical data

Supporting preclinical data

Expert Panel Review
- Policy
- Industry
- NGO
- Advocacy
- Regulators

After regulatory approval of a new drug

INTERNATIONAL ONLINE ENCYCLOPAEDIA

Sequence depository
- Routine release of high-quality clinical sequence data by accredited laboratories
- User-friendly interface to interrogate data

Forum
- New markers for drug resistance and treatment outcome are discussed and proposed

Diagnostic database
- Approved markers for automated WGS data interpretation

Review by expert panel

Scientific community

REFERENCE LABORATORY

Secondary surveillance and analysis

Phenotypic analysis

Patient

Doctor

Sequence data

Complex samples and QC

Treatment outcome

Enrichment/isolation of pathogens (molecular methods or culture) and WGS

Fully automated analysis to interpret results which are of immediate clinical relevance

Anonymised sequence data, New markers for drug resistance and virulence determinants

WHO, other governments
Gaps: Data capture and standardization

• Clinical data
  - Historically has been difficult to capture
  - Prior drug exposure(s) and clinical outcome (definitions)
  - Define standard reporting requirements (CDISC)

• Existing DST data
  - Phenotypic data (solid/liquid, direct/indirect, enzymatic)
  - Molecular data (LPA, Xpert)
  - Critical to define data quality (gold standard)

• Sequencing data
  - What to capture: SNPs, read depth, whole genome
  - Sequencing platforms, storage and bioinformatics?
  - Define standards for curation and base calling?
  - Define quality of sequencing data (quality score cutoffs)?

• Demographic data
  - Geographic location, gender, age...
Elements to consider

Intended use
- R&D (SNP discovery)
- Surveillance
- Biomarkers
- Clinical management
- Regulatory

Governance
- Curation
- Access rights
- Academic authorship
- IP
- Privacy

Database management
- Data standards
- Levels of access
- Cloud Infrastructure
- Data archive
- Data quality

Stakeholders
- Researchers/Clinicians
- Public Health Officials
- Policy makers
- Governments
- NGOs
- For profit/NFP

Sequencing
- NGS platforms
- Read depth
- Sanger sequencing
- Quality scores

Genetic-based tests
- Line Probe Assays
- Xpert MTB/RIF
- Spoligotyping

Clinical data
- Prior drug treatment regimens
- Coinfections
- metabolic, NCD, nutrition
- Patient outcomes

Phenotypic assays
- Solid media (LJ/Middlebrook)
- Nitrate-reduction
- Liquid (MGIT/VersaTREK/MODS)
- Phage-based
- Colorimetric

TB biorepository or specimen banks

Other TB databases

Global TB Database Consortium

Database management

A few questions to address

• What do we need to capture and why?
  - Immediate needs (SNPs, MICs, and clinical data for existing drugs)
  - Next steps (read depth for heteroresistance, new drugs, and WGS)
  - Future needs (mass spec, proteomic, transcriptomics, metabolomic...)

• How to ensure quality laboratory data?
  - Conventional DST assays (ISO accreditation, EQA, annotation)
  - Sequencing platforms (common metric for determining quality scores)
  - Automated curation (alignment, SNP and indel scan, mutation matrix)
  - How to assure consistency and when is human intervention needed

Proposed key outputs:
1. Consensus quality standards for DNA sequencing
2. Consensus for quality phenotypic data (PZA)
3. Best practices document for downstream analysis procedures

• Utilize CPTR and NDWG infrastructure to coordinate
  - Follow up discussion, conference calls and/or face-face meeting
  - Position paper
Acknowledgements

• Debra Hanna and Lindsay Lehman (C-Path)
• Jim Gallarda (BMGF)
• James Posey (US CDC)
• Richard Hafner (NIAID, NIH)

Meeting organizers

• John Ridderhof
• Daniela Cirillo
• Alessandra Varga
• Ruth McNerney
Sequence quality

Next generation WGS questions
- How do the different sequencing platforms compare?
- Agreement on a reference genome?
- Differences in bioinformatic software?
- How to identify mapping errors and artifacts that lead to false positive mutations?

Potential quality indicators
- Cross contamination quality checks
- Quality checks for:
  - NTB and mixed infection
  - Strand bias
  - Synonymous/non-synonymous
- Minimum coverage (>120 fold)
- Mapping quality score (>50)
- Base quality score (>30)
Additional Challenges

• Legal and policy issues
  - New Drugs: Industry collaborations
  - Existing Drugs: Country cooperation (India, China, former Soviet republics)

• How to ensure continued support?
  – Lessons learned: TB Dream, TDR, academic, surveillance, industry
  – User friendliness must be a priority
  – Bioinformatics and IT for different sequencing platforms
  – Maintenance and funding challenges (NIH, Welcome trust, MRC...)

• Link to other efforts
  – NIAID WGS contract (Broad Institute) and PATRIC database
  – Horizons 2020 (database tied to repository)