EBioMedicine 2 (2015) 1262–1265

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

In Focus

A Collaborative Approach for "ReSeq-ing" *Mycobacterium tuberculosis* Drug Resistance: Convergence for Drug and Diagnostic Developers

Marco Schito [a], David L. Dolinger [b]

[a] Critical Path Institute, 1730 E River Rd, Suite 200, Tucson, Arizona, United States
[b] FIND, Campus Biotech, Building B2, Level 0, 9, Chemin des Mines, 1202 Geneva, Switzerland

ARTICLE INFO

Article history:
Received 1 October 2015
Accepted 5 October 2015
Available online 9 October 2015

Keywords:
Tuberculosis
Drug resistance
Sequencing
Database

Since 2002 there has been a gradual, 1.3% per year, decrease in the incidence of tuberculosis (TB) worldwide. Though this is encouraging, this level of reduction will not meet the goal of the World Health Organization (WHO) to eliminate TB by 2050. In addition, the problem is only being compounded by the growing incidence of drug-resistant tuberculosis (Dye et al., 2013). Keys to the effective control of TB and the spread of drug resistant strains include the quick and accurate procurement of drug susceptibility information from patient samples that are infected with *Mycobacterium tuberculosis* complex (MTBC), linkage to appropriate treatment regimens, and follow-up to ensure cure (Wells et al., 2013). Phenotypic culture-based solutions, although the current "gold standard", are cumbersome, time-consuming, and consequently have a high percentage of loss to follow-up. Furthermore, culture-based solutions expose laboratory workers to potential risk of infection and certain drugs have assay reproducibility and accuracy concerns. Moreover, there are a multitude of phenotypic tests using different methods and growth media which result in discordant phenotypic results depending on the drug being tested. Detection of resistance conferring mutations by molecular methods is a promising alternative, but there is a paucity of information that correlates the infecting MTBC strain's genotype to phenotype and ultimately to patient outcomes (Noor et al., 2015).

To take better advantage of the present and emerging knowledge on genetic patterns encoding for drug resistance, there is a need to establish a sustainable data sharing platform and associated bioinformatic analysis tools. The platform will need to be able to house genotypic data, its related phenotypic data and associated metadata from individual isolates and be easily maintained, accessed and queried. The platform should contain the current wealth of knowledge that is located in disparate databases located in multiple research centers from around the world. But it will also be critical to validate the consistency and quality of the data added to the platform, and organize and maintain contemporaneous validated datasets. This will be especially important for data related to off-label use of repurposed drugs currently used to treat TB (Sotgiu and Migliori, 2015).

To prevent drug selection pressure from nurturing the development of resistance in treated populations, it will be imperative to align the roll out of new drug (Evangelopoulos and McHugh, 2015) and drug regimens (Fig. 1) with novel diagnostic tools to monitor drug resistance. Furthermore, drug developers need to revise their informed consent in these trials so that MTBC isolated from treated patients can be used in research to further dissect clinically relevant drug resistance mechanisms. The proper management of drug resistance is more cost effective than developing new anti-infective drugs and will secure the efficacy of these newly developed antimicrobials for the future.

The currently available genotypic and correlated phenotypic data concerning the resistance of MTBC have not been vetted or validated in a standardized manner, nor have the multiple data sets been merged. Moreover, many of the data sets are not curated (in terms of quality control) and may not be properly maintained. Clinical information (e.g., treatment outcome) is present only sporadically. Therefore, the first challenge lies in the development of a well-structured data platform that integrates high-quality sequencing data using a unified standardized pipeline from different research centers together with the relevant complementary information including associated drug susceptibility testing phenotypes, minimal inhibitor concentrations, resistance associated mutations, single nucleotide polymorphisms (SNPs), and additional clinical outcome data. While a limited set of genetic resistance markers for tuberculosis drugs has been identified, knowledge gaps still exist which impede the broader use of genomic information for the prediction of drug resistance (Salamon et al., 2015).

With the introduction of new drug combinations and regimens, and patients with potentially more complex resistance profiles, it is imperative to understand SNPs and genotypic information, and their correlation with phenotypic resistance, in order to allow for the implementation of correct, personalized therapies in a rapid, individual

E-mail address: mschito@c-path.org (M. Schito).

http://dx.doi.org/10.1016/j.ebiom.2015.10.008

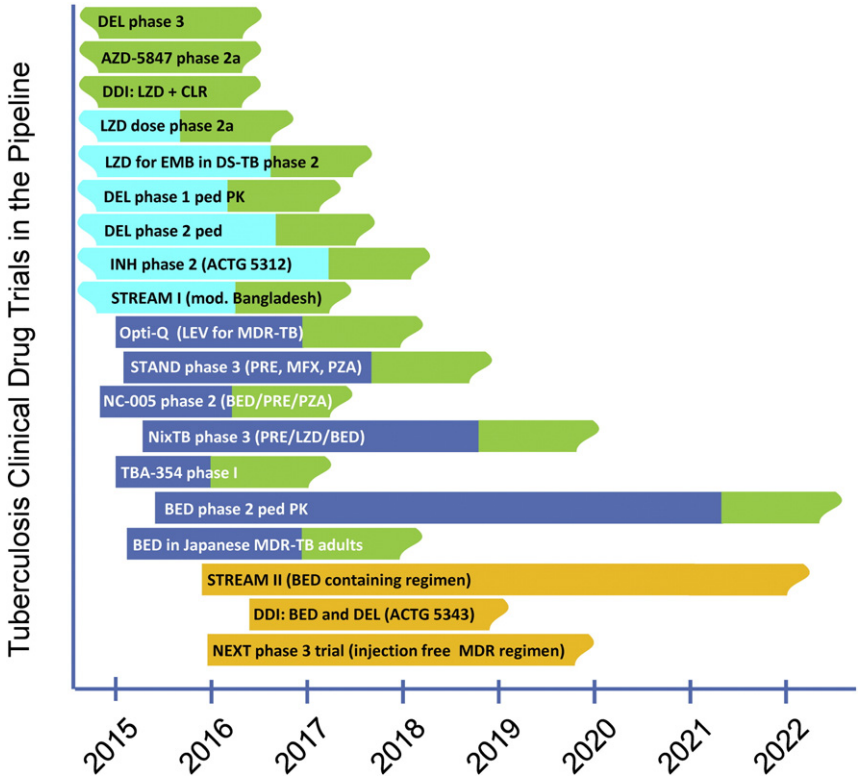

Fig. 1. Landscape of current and planned tuberculosis drug and new drug regimen clinical trials. A representation of various TB trials that are in analysis (green bars), began prior to 2015 (light blue bars), initiated in 2015 (dark blue bars), and not yet started (yellow bars). Abbreviations: DDI, drug–drug interaction; DS, drug sensitive; MDR, multi-drug resistant; ped, pediatric; PK, pharmacokinetic. Drug abbreviations: CLR, clarithromycin; DEL, delamanid; EMB, ethambutol; INH, isoniazid; LEV, levofloxacin; LZD, linezolid; MXF, moxifloxacin; PRE, pretomanid; PZA, pyrazidamide.

patient manner. While we don't yet completely understand the relationship and mechanisms of resistance in MTBC, we do have an understanding of the correlation between genotype and phenotypic resistance. Despite a wealth of current knowledge, new genotypic and phenotypic information is evolving which will provide higher confidence in the relationship between a specific SNP and phenotypic drug resistance. However, this information can only be effectively utilized if there is a centralized data platform available. Such a platform including a unified analysis pipeline must be vetted and accepted by the TB community, researchers, developers, clinicians, and patients to the point that it can become a sustainable platform.

New sequencing technologies are providing an ever increasing avalanche of genomic information at continuously lower cost. However, the use of this information is lagging for a number of reasons. Firstly, there is a lack of understanding of the clinical significance of the genomic data and the complex mutual interdependence of specific mutations. Secondly, some polymorphisms are either synonymous, i.e. unlikely associated with resistance, or are lineage specific markers. Finally, there are challenges in analyzing the data which include the lack of validated references and suitable validated interpretation tools. This is especially important when considering the underlying complex genetic interactions, like epistatic effects, for different mutations on drug resistance (Fenner et al., 2012). These challenges will further increase as the data creation continues to outpace the capabilities for data collection, organization, analysis and interpretation.

The ReSeqTB data sharing platform is a potential solution (Starks et al., 2015). The platform is being designed and developed by a unique collaboration between WHO, US CDC, New Diagnostics Working Group, Critical Path Institute, FIND and a number of academic TB experts. The goal of this initiative is to design, develop, populate and make a sustainable data sharing platform which will be used to aggregate genotypic, phenotypic and associated meta-data for MTBC. The initiative has been developed in a phased manner with the first set of objectives centered on the design and development of the platform and the expansion of an initial characterization of SNPs that have been observed in MTBC. Raw sequencing data obtained from contributors will be curated, standardized and aggregated through a unified pipeline. In exchange for their data, contributors will receive ongoing access to high-quality curated, annotated, globally aggregated datasets which can be utilized for database interrogation, pertinent statistical data correlations, and meaningful data outcomes and applications. As a further incentive, their contributions will result in individual and institutional researcher acknowledgement by way of credit within relevant publications.

Accelerated development of the proposed advanced bioinformatic tools for use with the ReSeqTB platform is possible due to the complementary skills of the partners in the initiative bringing together subject matter experts in data warehousing, scientific experts, and end users who will test and improve the tools in an iterative manner. Plans are also in development to allow researchers access to raw data to accelerate the development of new products and to improve case management in the clinical setting. The ReSeqTB platform can be utilized to minimize diagnostic costs, advance the research agenda and reduce testing times which could ultimately benefit patients, through advances in basic research (sequencing of resistance genes) to clinical research (identification of new drug targets), new diagnostic products (faster diagnosis of resistant bacteria and viruses), improved medical treatment (avoidance of antibiotic resistance due to inadequate drugs and dosages) and reduced health expenditures on a national level (avoidance of expensive 2nd and 3rd line drug treatments) (Table 1).

Initial results are already indicating new, but also more complex answers to the question of the genetic foundations of antibiotic resistance, as demonstrated by the discovery of additional putative genes and the indication of intergenic non-coding regions associated with drug resistance (Walker et al., 2015). The ReSeqTB platform will make it possible to better understand the underlying biology governing complex resistance profiles, to foster development of more powerful and targeted

Table 1
Stakeholder needs and benefit analysis.

Stakeholder	Interests	Needs	Benefits
Patient	• Proper, accurate and effective treatment of their MTBC infection	• Rapid development and deployment of assays for the accurate and sensitive detection of MTBC resistance • New and more effective therapeutic regimens for resistant MTBC and for decreasing the potential for developing resistance	• Faster and more accurate diagnosis of drug resistant MTBC • More appropriate and effective therapeutic regimens for drug resistant MTBC • More effective therapeutic regimens for non-resistant MTBC which decreases the potential for resistance
Individual academic data contributor	• Basic research in drug resistance mechanisms • Basic research in drug development	• Access to high quality, curated, annotated, globally aggregated data sets for research purposes • Ability to utilize the data sets for new investigations • Access to the results of raw data processed through the unified pipeline • Access to data analysis tools (long term deliverable)	• Virtual collaborative environment • Co-authorship on output which utilizes their data
Pharma co. data contributor	• Access and utilization of high quality, curated, annotated, globally aggregated data sets for research purposes	• Access to high quality, curated, annotated, globally aggregated data sets for research and development purposes • Access to the results of raw data processed through the unified pipeline • Ability to analyze sequence data from clinical trials • Access to data analysis tools (long term deliverable) • CDISC standardized datasets	• Ability to compare analyzed clinical trial data against high quality, curated, annotated, globally aggregated data sets and validated resistance associated mutation sets
DST developer end user	• Validated list of resistance associated mutations • Interpretation/meaning of resistance associated mutations	• Access to high quality, curated, annotated, globally aggregated data sets for research and development purposes • Access to the results of raw data processed through the unified pipeline • CDISC standardized datasets	• Curated database of MTBC genomic sequences • Validated list of resistance associated mutations with associated reference and supporting data
Research end user	• Access and utilization of high quality, curated, annotated, globally aggregated data sets for research purposes	• Ability in a controlled and validated manner to reanalyze MTBC sequence data • Ability to compare analysis data against a validated set of resistance associated mutations for MTBC • Ability to download the aggregated ReSeqTB Database • Access to data analysis tools for new investigations (long term deliverable)	• Consensus driven interpretation of MTBC sequence data • Ability to data mine an MTBC sequence database of high quality curated annotated globally aggregated data sets
Treatment advocacy group	• Patient treatment with the 'best' regimen possible based upon the patient's sequence data	• Validated resistance associated mutation list that is updated as often as possible and which develops the list for new drugs and drug regimens prior to the their general implementation	• Proactive database of validated resistance associated mutations (updated prior to release of new drugs and drug regimens)
Funding organizations	• Impact on funded projects	• Guidance information and materials to allow for proper assessment of proposals • Metrics for assisting in the measure of impact for funded proposals	• Validated, consensus driven database and analysis pipeline for funded proposals in MTBC sequencing projects
High burden and low burden country's ministries of health	• Tools for increasing the effectiveness of treatment in such a manner that improves overall treatment outcomes and decreases the economic impact of treatment	• Decreasing the rate of infection for MTBC and drug resistant MTBC • Tools for increasing the effectiveness of treatment • Active MTBC surveillance	• Potential for beneficial economic impact on the diagnosis and treatment of drug resistant MTBC • Validated list of resistance associated mutations with associated reference and supporting data • Increased impact with sequence based surveillance studies

drugs, to provide timely and personalized therapies and to contribute to current surveillance efforts. Ultimately, ReSeqTB will help to fight the further spread of drug resistant strains within a community by facilitating the potential for guided and personalized therapeutic regimens which cannot be achieved with presently available approaches.

Disclosure

The authors declared no conflicts of interest.

References

Dye, C., Glaziou, P., Floyd, K., Raviglione, M., 2013. Prospects for tuberculosis elimination. Annu. Rev. Public Health 34, 271–286.

Evangelopoulos, D., McHugh, T.D., 2015. Improving the tuberculosis drug development pipeline. Chem. Biol. Drug Des. (Mar 16) http://dx.doi.org/10.1111/cbdd.12549.

Fenner, L., Egger, M., Bodmer, T., Altpeter, E., Zwahlen, M., Jaton, K., Pfyffer, G.E., Borrell, S., Dubuis, O., Bruderer, T., Siegrist, H.H., Furrer, H., Calmy, A., Fehr, J., Stalder, J.M., Ninet, B., Böttger, E.C., Gagneux, S., Swiss HIV Cohort Study and the Swiss Molecular Epidemiology of Tuberculosis Study Group, 2012. Effect of mutation and genetic background on drug resistance in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. 56, 3047–3053.

Noor, K.M., Shephard, L., Bastian, I., 2015. Molecular diagnostics for tuberculosis. Pathology 47, 250–256.

Salamon, H., Yamaguchi, K.D., Cirillo, D.M., Miotto, P., Schito, M., Posey, J., Starks, A.M., Niemann, S., Alland, D., Hanna, D., Aviles, E., Perkins, M.D., Dolinger, D.L., 2015. Integration of published information into a resistance-associated mutation database for *Mycobacterium tuberculosis*. J. Infect. Dis. Suppl. 2, S50–S57.

Sotgiu, G., Migliori, G.B., 2015. Facing multi-drug resistant tuberculosis. Pulm Pharmacol Ther. 32, 144–148.

Starks, A.M., Avilés, E., Cirillo, D.M., Denkinger, C.M., Dolinger, D.L., Emerson, C.E., Gallarda, J., Hanna, D., Kim, P.S., Liwski, R., Miotto, P., Schito, M., Matteo, Z.M., 2015. Collaborative effort for a centralized worldwide tuberculosis relational sequencing data platform. Clin. Infect. Dis. 61 (Suppl. 3), S141–S146.

Walker, T.M., Kohl, T.A., Omar, S.V., Hedge, J., Del Ojo Elias, C., Bradley, P., Iqbal, Z., Feuerriegel, S., Niehaus, K.E., Wilson, D.J., Clifton, D.A., Kapatai, G., Ip, C.L., Bowden, R., Drobniewski, F.A., Allix-Béguec, C., Gaudin, C., Parkhill, J., Diel, R., Supply, P., Crook, D.W., Smith, E.G., Walker, A.S., Ismail, N., Niemann, S., Peto, T.E., Modernizing Medical Microbiology (MMM) Informatics Group, 2015. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. Lancet Infect. Dis. 15, 1193–1202.

Wells, W.A., Boehme, C.C., Cobelens, F.G., Daniels, C., Dowdy, D., Gardiner, E., Gheuens, J., Kim, P., Kimerling, M.E., Kreiswirth, B., Lienhardt, C., Mdluli, K., Pai, M., Perkins, M.D., Peter, T., Zignol, M., Zumla, A., Schito, M., 2013. Alignment of new tuberculosis drug regimens and drug susceptibility testing: a framework for action. Lancet Infect. Dis. 13 (5), 449–458.