Tuberculosis xxx (2015) 1–4

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: http://intl.elsevierhealth.com/journals/tube

Whole genome sequencing: A new paradigm in the surveillance and control of human tuberculosis

Seyed E. Hasnain [a, b, *], Ronan F. O'Toole [c], Sonam Grover [a], Nasreen Z. Ehtesham [d]

[a] Kusuma School of Biological Sciences, Indian Institute of Technology (IIT), Hauz Khas, New Delhi 110016, India
[b] Dr. Reddy's Institute of Life Sciences, University of Hyderabad Campus, Prof. C. R. Rao Road, Gachibowli, Hyderabad 500007, India
[c] Department of Clinical Microbiology, School of Medicine, Trinity College, Dublin, Ireland
[d] National Institute of Pathology, Safdarjung Hospital Campus, New Delhi 110029, India

ARTICLE INFO

Article history:
Received 3 October 2014
Accepted 22 December 2014

SUMMARY

Whole Genome Sequencing (WGS) is emerging as a very powerful tool for the management, outbreak analyses, surveillance and determining drug resistance of human infectious pathogens including *Mycobacterium tuberculosis* and MRSA. WGS can also discriminate relapse TB from re-infection and the resolution provided by WGS has no comparison to conventional technologies. With current cost coming down to <£70 per bacterial genome, WGS has emerged as an alternative to all the existing technologies put together. We discuss the advantage and disadvantages of WGS and whether it can become a point of care tool in not just developed countries but also in developing countries which have a huge TB burden. The likely utility of WGS for other pathogens and also in characterizing holobionts is also discussed.

1. Existing techniques for pathogen typing

Early laboratory medicine relied on phenotypic traits to identify infectious microorganisms. The advent of molecular biology made possible the detection of genetic markers that are unique to individual microbial species and trace the evolutionary dynamics and population structure of human pathogens [1,2]. Underpinning molecular epidemiology is the ability to differentiate individual strains, and hence, identify the source and route of transmission of an infection.

A wide range of techniques have been developed for the genotyping of microbes in clinical microbiology settings. In relation to *Mycobacterium tuberculosis* complex, commonly used techniques include Restriction Fragment Length Polymorphism (RFLP), Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), and Multiple Loci Variable-number tandem repeat Analysis (MLVA). The latter technique in particular, referred to as Mycobacterial Interspersed Repetitive Units (MIRU)-Variable Number Tandem Repeats (VNTR), has been considered one of the gold-standard techniques in the molecular typing of clinical isolates of *M. tuberculosis* complex [3–6].

All of the above techniques have been used to supplement epidemiological data and enable the determination of outbreaks of tuberculosis (TB), however, they possess a number of intrinsic limitations [7]. Many of the markers used are located in variable regions of the genome and therefore, do not necessarily provide an in-depth phylogeny of the infecting microorganism. Secondly, they generate little data from which to predict the phenotypic properties of a new strain, in particular, whether an isolate encodes resistance to first- or second-line drugs that are relied upon to treat TB. In addition, as a limited number of genetic markers are analyzed by RFLP, CRISPR or MLVA, the level of resolution can make it difficult to identify differences in isolates between contacts, and to accurately elucidate the direction of transmission. Presented with the constraints of existing mycobacterial genotyping techniques, it is perhaps not surprising that the arrival of low-cost whole genome sequencing (WGS) with a reduced turn-around time, has led to a rapid use of WGS in the study of *M. tuberculosis* complex. WGS was used for the first time with the twin purpose of ascertaining clustering of TB cases and quantification of active transmission, we present some of the highlights from recent applications of WGS to TB clinical microbiology and disease management.

(i) Distinguishing TB patient relapse and re-infection

* Corresponding author. Kusuma School of Biological Sciences, Indian Institute of Technology (IIT), Hauz Khas, New Delhi 110016, India. Tel.: +91 1126597522.
E-mail addresses: sehiitd@gmail.com (S.E. Hasnain), otooler3@tcd.ie (R.F. O'Toole), nzehtesham@gmail.com (N.Z. Ehtesham).

http://dx.doi.org/10.1016/j.tube.2014.12.007
1472-9792/

The ability to distinguish between relapse and re-infection was assessed in a recent ground breaking study by Bryant and co-workers [8]. Clinical trials for TB treatment are normally based on the supposition that when disease re-occurs, it is due to relapse. Isolates were taken by sputum sampling of patients enrolled in a randomized controlled TB treatment trial before treatment, and also at recurrence of disease [8]. 47 paired isolates were examined using 24-loci MIRU-VNTR and also for single-nucleotide polymorphisms (SNP) with WGS. While MIRU-VNTR analysis could detect re-infection with a distinct strain, it exhibited lower discriminatory capability with respect to WGS whereby a significant number of relapse TB cases were classified as re-infection by MIRU-VNTR. Furthermore, cases that were identified as mixed infections by WGS were classified as either relapse or re-infection by MIRU-VNTR. Therefore, to accurately monitor the success of a particular treatment regimen in a patient, or the prowess of a candidate drug in clinical trials, WGS fulfills an important function in being able to distinguish between relapse and re-infection.

TB recurrence poses a significant problem to the global control of TB. For example, of 306 HIV-positive TB patients who were deemed cured of TB in a study in India, recurrent TB occurred in 14% of patients [6]. In an earlier study involving 663 South African gold-mine workers who had an index episode of TB, 21% of HIV-negative men, and 34% of HIV-positive men developed recurrent TB [5]. Application of WGS to such studies in future should improve our understanding of TB relapse and re-infection in high-risk populations. The very high resolution is particularly useful in categorically defining recent transmission events [9].

(ii) Mapping community TB outbreaks

The detection of intra-strain microevolution by WGS is finding applications in the investigation of community outbreaks of TB for determining the direction of spread between cases, and the identification of so-called “super-spreaders”. Walker and co-workers estimated a rate of change in DNA sequences at 0.5 SNPs per genome per year (95% CI 0.3–0.7) by longitudinal sampling of patients and households in the UK Midlands and WGS of the resulting 390 isolates [10]. 96% of paired isolates from individual patients or households were separated by 5 or fewer SNPs. Hence, the authors used an upper threshold of 5 SNPs to predict epidemiological linkage consistent with transmission. This WGS-based analysis enabled them to identify cryptic outbreaks in geographically separated patients for which there were no recorded epidemiological links [10]. It also facilitated the identification of the likely direction of transmission between patients and the detection of so-called “super-spreaders”. Other researchers have estimated similar genome evolutionary rates of 0.4 mutations per genome per year [11] and 0.3 mutations per genome per year [12]. Using their rate, Roetzer and co-workers established that the Hamburg clone responsible for a TB outbreak in Germany emerged between 1993 and 1997, shortly before the outbreak was detected by epidemiological surveillance [11]. In high TB burden settings, WGS has been shown to be useful in tracing *M.tb* transmission [9,13]. In another recent study it was shown that most of the TB patients residing in Oxfordshire were cases of reactivation of latent infection in acquired while they were living in high residence countries before immigrating to UK [10]. A TB outbreak over 21 years was traced in real time and at high resolution by a combination of WGS and genome network construction [14].

(iii) Detection of anti-tubercular drug resistance

While the existing techniques of RFLP, CRISPR (spoligotyping) or MLVA (MIRU-VNTR) can distinguish *M. tuberculosis* complex strains and identify clusters, they provide little information on the phenotypic properties coded by a previously-uncharacterized isolate. Hence, determination of an isolate's drug susceptibility is still heavily reliant upon phenotypic assays which take several weeks to perform due to the slow growth rate of *M. tuberculosis* complex. Improvement in diagnostic technologies is needed to enable early identification of drug-susceptible, multi-drug resistant (MDR), and extensively-drug resistant (XDR) forms of TB. This would in turn facilitate the optimal selection of anti-TB drugs and the minimization of further transmission of resistant strains.

In a recent study, Köser and colleagues [15] used WGS to investigate a case of XDR-TB. DNA was extracted directly from a 3-day positive MGIT (mycobacterial growth indicator tube) culture and sequenced using an Illumina® MiSeq bench-top instrument. Although mixed infection had not been detected using MIRU-VNTR, WGS identified two distantly related Beijing strains of *M. tuberculosis* complex in the patient's sputum. Furthermore, WGS identified mutations linked with resistance to a number of TB drugs that were not phenotypically tested for at the reference laboratory [15].

While detection of mutations in genes associated with resistance is facilitated by WGS for multiple anti-TB drugs, it is important that this sensitivity is appropriately balanced with high confidence levels of specificity. For this reason, Rodwell and co-workers collected 416 patient isolates from regions with a high burden of MDR/XDR-TB and subjected them to WGS analysis in parallel with standardized phenotypic drug-susceptibility testing (DST) [16]. This resulted in a description of dominant SNPs that have high specificity, $\geq$95%, for resistance to key TB drugs rifampicin, isoniazid, fluoroquinolones, and aminoglycosides. This extends the power of WGS in relation to DST and provides a template for the use of precise predictors of drug resistance in clinical decision-making on the formulation of an anti-TB therapy for a patient. Additionally, WGS enables investigation on evolution, transmission and prevalence of drug resistance within a defined population [17].

(iv) Present limitations

Although the potential applications of WGS in clinical microbiology are considerable, a number of limitations associated with the technology need to be acknowledged. For example, WGS is technology intensive and therefore, requires substantial investment in next-generation sequencing equipment such as Illumina® instrumentation. This upfront cost may fall outside the financial resources of many diagnostic facilities in not so industrialized countries, and more so in developing countries. The cost of laboratory consumables, while coming down, are not quite yet at the point where WGS is more cost effective than other techniques used for molecular typing of *M. tuberculosis* patient isolates. Where equipment and consumable procurement is possible, the nature of the laboratory work associated with WGS, means that a high degree of training of laboratory personnel is required to process samples for sequencing. On a more fundamental note the fact that *M.tb* isolates are not at all homogenous population but comprise of population intrinsically different between themselves in terms of SNP profile poses real problem. The WGS read out is for the predominant strains which may or may not be relevant during patient to patient transmission. Another issue associated with WGS is the massive quantity of data that it produces and that must be analyzed therefore posing a bioinformatic challenge [18]. Analysis can therefore be time-consuming and prone to error unless highly-experienced bioinformaticians are employed. The presence of repetitive gene sequences in the *M. tuberculosis* genome, in particular the mostly hypothetical PE/PPE/PGRS gene family [19], present nowhere in the living world except in the genus Mycobacterium, also presents challenges to the attainment of complete genome sequence

Please cite this article in press as: Hasnain SE, et al., Whole genome sequencing: A new paradigm in the surveillance and control of human tuberculosis, Tuberculosis (2015), http://dx.doi.org/10.1016/j.tube.2014.12.007