EBioMedicine 36 (2018) 14–15

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Commentary

Precision tuberculosis control by genome sequencing: Benefit and challenges of a new standard

Philip Supply

Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 8204 - CIIL - Centre d'Infection et d'Immunité de Lille, F-59000 Lille, France

Progress in the control and prevention of tuberculosis (TB) transmission will be necessary for reaching the goal set by the World Health Organization to end the global TB as a public health problem, at least in low-incidence countries by 2050 [1]. In these countries, the disease is more concentrated in vulnerable and hard-to-reach risk groups, such as homeless and migrants. This complexifies contact investigation around TB cases, representing a central component of outbreak control and prevention [2]. Genetic pairing between *Mycobacterium tuberculosis* isolates by a standardized 24-marker-based genotyping (MIRU-VNTR) is therefore widely used to guide investigation towards probable transmission links [3]. However, this genotyping overestimates clusters of ongoing transmission, and is being supplanted by whole genome sequencing (WGS), which is increasingly affordable and offering ultimate resolution for precisely delineating TB transmission [4].

In an article in *EBioMedicine*, Wyllie and colleagues quantify for the first time at population scale the public health benefit that can be expected with this replacement [5]. To this end, they systematically compared MIRU-VNTR pairing and sharing of various epidemiological risk factors between TB cases versus relatedness by WGS, in an unselected patient population of the English Midlands. As a central postulate, close genome-wide relatedness of 5 or less single nucleotide variants (SNVs) between isolates was used as a reference to define recent transmission, against which MIRU-VNTR-paired isolates and shared risk factors were matched. This postulate is supported by levels of SNV variation generally observed in *M. tuberculosis* genomes in recent transmission chains, in epidemiologically similar settings [4]. The foremost finding of this evaluation is that, overall, any pair of isolates matched by identical MIRU-VNTR types positively predicted only between ~15% and ~50% of close WGS-based relatedness suggestive of recent transmission links, depending on whether shared risk factors (excluding simple co-residence) were not or were recorded between patients. The predictive values of MIRU-VNTR linked pairs for possible transmission, as defined using a SNV cutoff, varied substantially among lineages of *M. tuberculosis* strains: it was higher in lineages associated with autochthonous patients, but lower in lineages associated with foreign born and recently immigrated patients, consistent with previous results [6]. From this, it follows that switching to WGS for detection of TB transmission clusters could result in a two- to seven-fold reduction in the number of potential links requiring in-depth epidemiological investigation, at least in demographically similar, low-burden settings.

In addition to confirming the superior precision of WGS at unprecedented scale, this calculated gain in epidemiological investigation efforts will further motivate the transition to this technology as a new standard for molecular-guided control of TB. This shift, ongoing in many (inter)national public health institutions, including in some high TB incidence settings, is also encouraged by the parallel effectiveness of genome sequencing for surveillance of drug resistant TB [7].

However, important definition and standardization challenges of key parameters remain to be met for optimal exploitation. Especially tricky is the need to define/adjust appropriate SNV breakpoints to assess transmission, e.g. depending upon possibly variable intensities of transmission and heterogeneity of strain populations prevailing among settings [4]. To re-validate their initially selected breakpoint in their study population, Wyllie and colleagues exploited the fact that spatial proximity to another case is a risk factor for transmission: They thus verified that the positive association between geographic proximity and close genomic relatedness vanished at SNV distances exceeding their breakpoint, as expected if transmission becomes unlikely beyond this limit. This quantitative validation approach may prove valuable elsewhere.

Yet, in the overall, shared risk factors as collected by the Enhanced Tuberculosis Surveillance system were strikingly found only in a minority of the cases assumed to be linked by WGS. One probable contributing factor is the under-detection of a substantial number of casual epidemiological links, which are difficult to spot by conventional epidemiological tracing. More dynamic cross-talk between rapid WGS and epidemiological tracing compartments in the TB control systems may help better uncover such difficult links. Overlay of anonymized contact information collected via mobile phone technologies on WGS results might represent an even more promising and futuristic prospect, although ascertainment and acceptability issues in at-risk populations must be considered.

Another potential explanation is that even close genome-wide relatedness partly overestimated recent transmission in this study. This might reflect some concomitant cases of reactivations of remote infections by a very same, once endemic strain. However, genome mutation rates might also be slower during certain transmission events, e.g. depending upon *M. tuberculosis* genetic backgrounds or the dynamic of transmission, implying that a fixed SNV breakpoint (5 here) to ascertain transmission may not universally apply even within a population. More research is needed to address these points.

DOI of original article: https://doi.org/10.1016/j.ebiom.2018.07.019.

E-mail address: philip.supply@ibl.cnrs.fr.

https://doi.org/10.1016/j.ebiom.2018.08.052

Other challenges concern the heterogeneity of the sequencing procedures and specifications of bioinformatics pipelines in use, making determinations of close genomic relatedness in part pipeline- and dataset-dependent. In particular, filtering of some (presumably) low confidence SNV variation, such as one newly used by Wyllie and colleagues [5,8], is not uniformly applied internationally. This currently limits direct comparability of close genome-wide relatedness, as illustrated by a recent international WGS investigation of a cross-border outbreak of multidrug resistant TB among migrants arriving in Europe. Initial tracking of outbreak cases among seven European countries was based on a common phenotypic drug resistance pattern, a shared MIRU-VNTR profile or shared drug resistance mutations. While most isolates had WGS in their country of diagnosis, the analysis of WGS data was centralized in one center to globally confirm the patients' links and properly delineate the outbreak [9].

Several large-scale projects are ongoing to set guidelines/standards and ensure better (inter)national integration of WGS-based systems [10]. The obvious obstacle of the implementation in low resource, high-incidence settings is also being tackled e.g. by development of simpler/more portable genome sequencing approaches. These collective efforts will allow to fully exploit the hoped-for transformative impact of this technology on TB control.

Disclosure

I am a consultant for Genoscreen.

References

[1] Lonnroth K, Migliori GB, Abubakar I, et al. Towards tuberculosis elimination: an action framework for low-incidence countries. Eur Respir J 2015;45(4):928–52.

[2] Erkens CG, Kamphorst M, Abubakar I, et al. Tuberculosis contact investigation in low prevalence countries: a European consensus. Eur Respir J 2010;36(4):925–49.

[3] Supply P, Allix C, Lesjean S, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. J Clin Microbiol 2006;44(12):4498–510.

[4] Walker TM, Ip CL, Harrell RH, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 2013;13(2):137–46.

[5] Wyllie D, Davidson J, Smith EG, et al. A quantitative evaluation of MIRU-VNTR typing against whole-genome sequencing for identifying Mycobacterium tuberculosis transmission: a prospective observational cohort study. EBioMedicine 2018;34:122–30.

[6] Stucki D, Ballif M, Egger M, et al. Standard genotyping overestimates transmission of Mycobacterium tuberculosis among immigrants in a low-incidence country. J Clin Microbiol 2016;54(7):1862–70.

[7] Zignol M, Cabibbe AM, Dean AS, et al. Genetic sequencing for surveillance of drug resistance in tuberculosis in highly endemic countries: a multi-country population-based surveillance study. Lancet Infect Dis 2018;18(6):675–83.

[8] Wyllie DH, Sanderson N, Myers R, et al. Control of artifactual variation in reported intersample relatedness during clinical use of a Mycobacterium tuberculosis sequencing pipeline. J Clin Microbiol 2018;56(8).

[9] Walker TM, Merker M, Knoblauch AM, et al. A cluster of multidrug-resistant Mycobacterium tuberculosis among patients arriving in Europe from the Horn of Africa: a molecular epidemiological study. Lancet Infect Dis 2018;18(4):431–40.

[10] Tagliani E, Cirillo DM, Kodmon C, van der Werf MJ, Consortium EU. EUSeqMyTB to set standards and build capacity for whole genome sequencing for tuberculosis in the EU. Lancet Infect Dis 2018;18(4):377.