EBioMedicine xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

A Quantitative Evaluation of MIRU-VNTR Typing Against Whole-Genome Sequencing for Identifying *Mycobacterium tuberculosis* Transmission: A Prospective Observational Cohort Study

David H. Wyllie [a,b,c,*], Jennifer A. Davidson [d], E. Grace Smith [e], Priti Rathod [e], Derrick W. Crook [a,c], Tim E.A. Peto [a,c], Esther Robinson [e], Tim Walker [a], Colin Campbell [d]

[a] *Nuffield Department of Medicine, University of Oxford, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK*
[b] *Public Health England Academic Collaborating Centre, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK*
[c] *The National Institute for Health Research, Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, UK*
[d] *Tuberculosis Section, National Infection Service, Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK*
[e] *Public Health England National Regional Mycobacteriology Laboratory North and Midlands, Heartlands Hospital, Birmingham BS9 5SS*

ARTICLE INFO

Article history:
Received 15 June 2018
Received in revised form 13 July 2018
Accepted 15 July 2018
Available online xxxx

Keywords:
Mycobacterium tuberculosis
Topic:
MIRU-VNTR
Single nucleotide variation
Outbreak investigation
Molecular epidemiology
Research in context

ABSTRACT

Background: Mycobacterial Interspersed Repetitive Unit-Variable Number Tandem Repeat (MIRU-VNTR) typing is widely used in high-income countries to determine *Mycobacterium tuberculosis* relatedness. Whole-genome sequencing (WGS) is known to deliver greater specificity, but no quantitative prospective comparison has yet been undertaken.
Methods: We studied isolates from the English Midlands, sampled consecutively between 1 January 2012 and 31 December 2015. In addition to routinely performed MIRU-VNTR typing, DNA was extracted from liquid cultures and sequenced using Illumina technology. Demographic and epidemiological data for the relevant patients were extracted from the Enhanced Tuberculosis Surveillance system run by Public Health England. Closely related samples, defined using a threshold of five single nucleotide variants (SNVs), were compared to samples with identical MIRU-VNTR profiles, to samples from individuals with shared epidemiological risk factors, and to those with both characteristics.
Findings: 1999 patients were identified for whom at least one *M. tuberculosis* isolate had been MIRU-VNTR typed and sequenced. Comparing epidemiological risk factors with close genetic relatedness, only co-residence had a positive predictive value of over 5%. Excluding co-resident individuals, 18.6% of patients with identical MIRU-VNTR profiles were within 5 SNVs. Where patients also shared social risk factors and ethnic group, this rose to 48%. Only 8% of MIRU-VNTR linked pairs in lineage 1 were within 5 SNV, compared to 31% in lineage 4.
Interpretation: In the setting studied, this molecular epidemiological study shows MIRU-VNTR typing and epidemiological risk factors are poorly predictive of close genomic relatedness, assessed by SNV. MIRU-VNTR performance varies markedly by lineage.
Funding: Public Health England, Health Innovation Challenge Fund, NIHR Health Protection Research Unit Oxford, NIHR Oxford Biomedical Research Centre.

Evidence Before This Study

We searched Pubmed using the search terms 'whole genome sequencing' and 'MIRU-VNTR' and 'tuberculosis' for English language articles published up to December 21st, 2017. Multiple studies have shown that most pairwise genomic comparisons will be within five SNVs when direct transmission has occurred from one individual to another. Both outbreak studies and population studies have demonstrated how whole-genome sequencing generates smaller clusters than MIRU-VNTR typing, and how sequence data allows for differentiation of isolates within a cluster. However, no systematic comparison of MIRU-VNTR typing vs. WGS has however been published. The degree to which WGS provides more specific results, and the degree to which it is likely to be more cost effective, therefore remains uncertain.

Added Value Of This Study

This study seeks to quantify the predictive value of identical MIRU-VNTR profiles, and of overlapping demographic and epidemiological data, for close genomic relatedness in a cosmopolitan setting. Importantly, it demonstrates that in our setting MIRU-VNTR-based clustering

* Corresponding author.
E-mail address: david.wyllie@ndm.ox.ac.uk (D.H. Wyllie).

https://doi.org/10.1016/j.ebiom.2018.07.019
/

predicts genomic relatedness differently depending on *M. tuberculosis* lineage. This is compatible with previous reports of poor discrimination by MIRU-VNTR in lineage 2 (Beijing), but is not restricted to lineage 2, and is likely to be generalizable to other settings. Our results provide an explanation as to why MIRU-VNTR typing was not cost effective when implemented in England, and indicate that WGS may perform substantially better.

Implications of All the Available Evidence

Whilst it is generally accepted that WGS provides more informative results than MIRU-VNTR typing, the latter is still practiced widely under the belief that it remains a helpful tool for public health investigations. This study shows that whilst differing MIRU-VNTR profiles help exclude close genomic relatedness, matching profiles rarely predict such relatedness. Having quantified its predictive value at a population level, this study should hasten the transition from MIRU-VNTR typing to WGS in other settings similar to ours.

1. Introduction

In 2016 there were 5664 notified cases of tuberculosis in the England, with an incidence of 10.2 per 100,000 population [1]. Despite a steady fall in incidence since its peak early this decade, this remains the highest rate in western Europe, outside of the Iberian peninsula [2]. This decline has occurred across almost all population groups with only a third due to decreases in the numbers of migrants from high TB burden countries. Despite decreases in TB rates, domestic transmission is still likely to be contributing to current case loads [3].

Rapid detection of *Mycobacterium tuberculosis* transmission should offer enhanced opportunities for disease control [4, 5]. In England, as in many high-income countries, tuberculosis transmission has been identified with the help of Mycobacterial Interspersed Repetitive Unit-Variable Number Tandem Repeat (MIRU-VNTR) typing, which clusters cultured isolates on the basis of their molecular fingerprints [6, 7]. A recent post-deployment evaluation of the MIRU-VNTR-based surveillance programme in England has however questioned the cost-effectiveness of this approach [8].

Since 2015, Public Health England has been undertaking a phased introduction of routine whole genome sequencing (WGS) for all mycobacterial cultures [9]. This has meant the relatedness of isolates could be simultaneously compared using both single nucleotide variants (SNV) and by MIRU-VNTR typing, and has provided a novel opportunity to compare the added value of whole genome sequencing ([10–15]; Table 1) in an unselected population, at scale. This approach contrasts with recent studies in which samples from diverse geographic locations were selected by lineage, with selected subsets being characterised by both SNV and MIRU-VNTR [16, 17]. Analysis of unselected samples, as practiced here, can be used to investigate reports that MIRU-VNTR typing differentiates parts of Lineage 2 [16] [18], as other lineages [19], poorly.

Here we estimate what proportion of *M. tuberculosis* isolates from a cosmopolitan area of central England that are linked by MIRU-VNTR typing, or have associated epidemiological risk factors, are closely genomically related. In this work, we use SNV as a metric of close genetic similarity; although other kinds of variation, including insertions and deletions (indels) exist [20], here we chose to use SNV, for which cutoffs reflecting close genetic relatedness have been derived a in a range of populations [21], and for which the clock rate has been heavily studied [21], including external calibration against historical events [16].

Table 1
Previous studies including both MIRU-VNTR and SNV analysis of *M. tuberculosis*.

Samples	Comment	Reference
36 archived Manila strain isolates	SNV analysis revealed variation not demonstrated by MIRU-VNTR.	10
390 retrospective isolates from the English Midlands	Genetic heterogeneity within MIRU-VNTR clusters demonstrated. 5 and 12 SNV proposed as potential cut offs for epidemiological relatedness.	11
199 epidemiologically linked cases sequenced retrospectively	Relationship with MIRU-VNTR profile was not addressed	37
36 isolates from an outbreak	SNV analysis revealed variation not demonstrated by MIRU-VNTR.	38
50 cases from an outbreak	SNV analysis revealed variation not demonstrated by MIRU-VNTR.	12
1000 isolate sample of 2248. Representative of Russian population studied, plus 28 diverse sequences	Relationship with MIRU-VNTR profile was not addressed. Multiple sub-lineages observed within Lineage 4 (Euro-American).	39
69 cases from an outbreak defined by a SNV	SNV analysis revealed variation not demonstrated by MIRU-VNTR.	13
86 cases from an outbreak	SNV analysis revealed variation not demonstrated by MIRU-VNTR.	14
90 cases belonging to 35 MIRU-VNTR clusters	MIRU-VNTR performance overestimated transmission particularly in immigrants infected with closely related strains	15
4987 lineage 2 samples representative of global diversity studied by MIRU-VNTR	110 specimen sample was sequenced by next-generation sequencing. MIRU-VNTR poorly defined some branches of the lineage 2 phylogeny	16
Paired isolates from 390 patient selected due to possible emergence of drug resistance	SNV analysis as well as MIRU-VNTR profiling used to confirm or exclude re-infection	40

2. Methods

2.1. Samples Studied for Comparison of MIRU-VNTR With SNVs

Consecutive *M. tuberculosis* isolates from the Public Health England Centre for Regional Mycobacteriology Laboratory, Birmingham between 1 January 2012 and 31 December 2015 were included in the study. This corresponds to the period when both MIRU-VNTR and SNV analysis were both performed. This laboratory serves a large catchment of approximately 12 million persons in the English Midlands, a region which includes high, medium (40–150 cases per 100,000 population), and low TB incidence areas. After exclusions, described in Results, 1999 isolates each isolated from a single patient, were studied.

2.2. Identification and MIRU-VNTR Typing

Clinical samples were grown in Mycobacterial Growth Indicator tubes (MGIT) (Becton Dickinson, New Jersey, USA), and *M. tuberculosis* was identified using Ziehl-Neelsen staining, followed by nucleic acid amplification and hybridisation using Genotype Mycobacterium CM hybridisation tests (Hain LifeScience, Nehren, Germany). 24-locus MIRU-VNTR typing [6, 22] was performed on the first isolate from each patient in each calendar year using non-denaturing HPLC (WAVE microbial analysis system) as described [23]. This assay demonstrated complete concordance with gel based fragment size analysis during the validation study in 2004 [23]. A detailed verification study, performed in 2014, indicated that assay performance had not changed substantially relative to the validation study (Supplementary Data 1). Throughout use, the assay was subject to internal and external quality control.

2.3. Laboratory and Bioinformatic Processing

This was carried out as described [11]. Nucleic acid was extracted from 1·7 ml of MGIT culture as described [9]. Illumina 150 bp paired