

**REVIEW**

# Translating genomics research into control of tuberculosis: lessons learned and future prospects

Digby F Warner<sup>1,2,3</sup> and Valerie Mizrahi<sup>1,2,3\*</sup>

## Abstract

Genomics research has enabled crucial insights into the adaptive evolution of *Mycobacterium tuberculosis* as an obligate human pathogen. Here, we highlight major recent advances and evaluate the potential for genomics approaches to inform tuberculosis control efforts in high-burden settings.

## Introduction

Tuberculosis (TB) is a leading cause of death as a result of an infectious bacterial agent, claiming 1.4 million lives each year [1]. With an estimated global burden of 8.7 million incident cases per annum, TB remains a major public health threat. In high-burden regions such as sub-Saharan Africa, the TB epidemic is exacerbated by co-morbidities, including HIV and diabetes, as well as demographic, socioeconomic, and programmatic factors [2]. The magnitude of the TB problem has been further amplified by the evolution and global spread of strains of *Mycobacterium tuberculosis* that are resistant to conventional first- and second-line antitubercular drugs. Of particular concern, drug resistance is worsening, having progressed from multi-drug resistant (MDR), to extensively drug resistant (XDR), to 'functionally untreatable' [3] TB - that is, disease for which no therapeutic options remain. This progression has led to calls for 'visionary political leadership' [4] and 'increased funding to sustain global control efforts, research and advocacy' [3]. In order to reach the aspirational goal of global TB elimination by 2050, TB incidence will need to be reduced by approximately 16% each year for the next 40 years. In

spite of recent gains in the battle against TB, the current rate of decline in TB incidence of 2% per annum falls far short of this target [5]. This alarming situation underscores the urgent need for new tools to control this devastating disease.

Fundamental TB research poses very specific practical and financial challenges. As an infectious pathogen, *M. tuberculosis* can only be manipulated in purpose-built biosafety level 3 containment laboratories by specialist personnel. The construction and maintenance of such facilities requires significant financial investment; moreover, the running costs necessary to ensure continued compliance with the stringent safety regulations are high, and are incurred in addition to standard laboratory operating expenses. From a practical perspective, *M. tuberculosis* is an intractable experimental subject: the bacillus is notorious for its slow growth rate *in vitro* and for its tendency to form aggregates in liquid media. As a result, experiments are technically demanding, long in duration and prone to contamination. The combined effect, therefore, is that the achievement of definitive results can be very slow.

Even more challenging are the scientific problems posed by the natural lifecycle of *M. tuberculosis* as an obligate human pathogen. By definition, all experiments conducted outside infected individuals - whether *in vitro* or *in vivo* - are performed in model systems that have varying capacities to recapitulate specific aspects of the host-pathogen interaction. Although advances in experimental mycobacteriology have provided key insights into the metabolic and regulatory pathways that are critical for bacillary survival and pathogenesis, it remains extremely difficult to determine the precise physiological status of tubercle bacilli during different stages of infection and in discrete anatomical and cellular (micro)environments. As noted elsewhere [6], an important consequence is that direct investigations of mycobacterial function in the context of the complete biological system - the *M. tuberculosis*-infected host - remain rare. In turn, this means that the barriers to translating the

\* Correspondence: valerie.mizrahi@uct.ac.za

<sup>1</sup>MRC/NHLS/UCT Molecular Mycobacteriology Research Unit, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, Cape Town 7925, South Africa

<sup>2</sup>DST/NRF Centre of Excellence for Biomedical TB Research, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Anzio Road, Observatory, Cape Town 7925, South Africa

Full list of author information is available at the end of the article

observations from basic research into practical outcomes are sizeable.

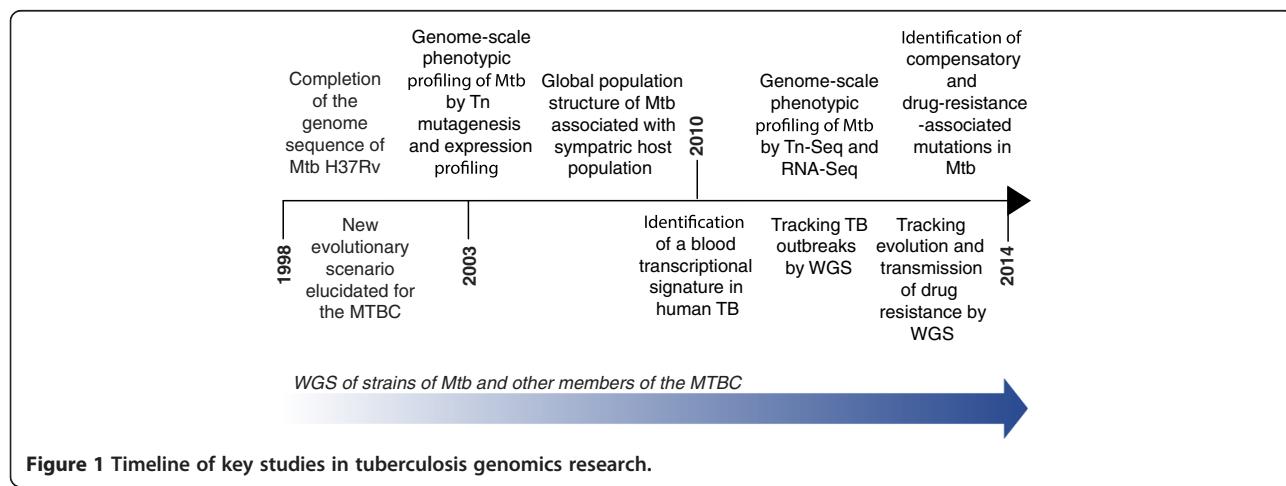
The application of genomics and other 'omics technologies in developing a systems biology of TB is central to global efforts towards the development of new vaccines, diagnostics and drugs for TB. The landmark publication in 1998 by Stewart Cole and colleagues [7] of the first genome sequence of a strain of *M. tuberculosis* ushered in a new era in TB research in which genome-scale studies have provided crucial insights into the ancient and modern evolutionary history of *M. tuberculosis*, the genomics of drug resistance, the biology of *M. tuberculosis* as an intracellular pathogen, and the host response to infection with this organism (Figure 1). In this article, we highlight the major advances in TB research that have been enabled by the genomics revolution. We then identify key areas of research and development that will be required in order to harness the full potential of genomics approaches for the control of TB in endemic regions, discuss some of the major challenges and obstacles that will need to be addressed and overcome in this endeavor, and conclude by considering the implications of the lessons learned from TB in the context of other infectious diseases.

### The evolutionary history of *M. tuberculosis*

*M. tuberculosis* is one member of the *M. tuberculosis* complex (MTBC), a collection of phylogenetically linked organisms comprising eight closely related lineages [8] and the outlying *M. canettii* group, in which the so-called 'smooth tubercle bacilli' are situated [9]. *M. tuberculosis sensu stricto* from lineages L1 to L4 and L7, together with the *Mycobacterium africanum* lineages L5 and L6, are human-adapted, whereas the L8 lineage - which includes *Mycobacterium bovis* and the TB vaccine strain, BCG (Bacille Calmette Guérin) - contains the animal-adapted pathogens. The recent discovery of chimpanzee and mongoose bacilli [10,11] suggests, however, that there might be

much greater diversity within the MTBC. In turn, this implies that existing scenarios for the evolution of the human- and animal-adapted strains might be overly simplistic, and limited by the availability of isolates, especially from wild mammals [11]. Defining the point in time, as opposed to the phylogenetic position, at which MTBC strains originated from a last common ancestor has proven very difficult [8,12,13]; nevertheless, both comparative genomics and bioarcheological evidence support the extended co-evolution of *M. tuberculosis* with its obligate host [14]. In turn, this implies the evolution of a conserved host-pathogen interaction that enables repetitive cycles of infection, disease, and transmission while accommodating bacillary adaptation to major human demographic shifts. Although not conclusive, evidence of selective pressure on specific mycobacterial antigens provides some support for this idea [15], as does the observation that diverse *M. tuberculosis* strains engage a core transcriptional response following macrophage infection, while exhibiting hallmarks of lineage-specific adaptation to geographically varied host populations [16]. Notably, the interaction between a particular locally adapted *M. tuberculosis* strain and its corresponding geographically matched host appears to depend on a functional immune response: these sympatric interactions are disrupted by HIV co-infection [17].

Unlike most other bacterial pathogens, a defining characteristic of *M. tuberculosis* is its reliance on chromosomal rearrangements and mutations as drivers of genomic evolution [14]. Horizontal gene transfer (HGT) certainly played an important role in the evolution of *M. tuberculosis* as a human pathogen [14,18,19]; however, despite the proposal that ongoing recombination provides a source of genetic variation [20], there is very little evidence in support of a role for HGT in the modern evolution of this organism [21]. This feature is likely to result from the ecological isolation of the bacillus as an obligate pathogen that primarily targets the host pulmonary and lymphatic



**Figure 1** Timeline of key studies in tuberculosis genomics research.

system [22], as well as from the severe bottlenecks imposed by aerosol-dependent transmission from infectious individual to naïve recipient [23].

### Insights from diversity between lineages of the MTBC

As noted above, the MTBC comprises eight closely related lineages [8] which can be distinguished according to a lineage-defining single nucleotide polymorphism (SNP) 'barcode' [24]. Until very recently, the functional consequences of almost all of the differentiating SNPs remained completely unexplored. In an important study illustrating the power of integrating 'omics with mycobacterial genetics and chemical biology in experimental models of TB infection, Christophe Guilhot, Roland Brosch and colleagues demonstrated that SNPs which are conserved in animal-adapted and *M. africanum* L6 strains are less transmissible and virulent in humans than *M. tuberculosis sensu stricto* [25]. Guided by insights from comparative genomics, these researchers homed in on three separate SNPs that map to the promoter region of *phoP* and codon 71 of *phoR*, genes which encode a two-component system previously implicated in the virulence and immunogenicity of *M. tuberculosis*. This system regulates the synthesis and export of virulence factors that include the major secreted antigen, ESAT-6, as well as polyacyltrehalose (PAT) lipids and sulfolipids (SLs). By transferring any of three alleles - *M. tuberculosis phoPR*, *M. bovis phoPR*, or a chimeric *phoPR* allele in which the *phoP* (promoter) and *phoR* (coding region) SNPs were split - into a *phoPR* null mutant of *M. tuberculosis*, the authors demonstrated that the *M. bovis phoR* allele is associated with impaired expression of the PhoPR regulon. The *M. bovis phoPR* allele was also found to impact negatively on mycobacterial virulence in human macrophage and mouse models of infection.

Armed with these data associating genotype with phenotype, the authors then set out to characterize the PhoPR system in a set of wild-type animal-adapted and *M. africanum* L6 strains, as well as in matched derivatives harboring the *M. tuberculosis phoPR* allele.

The levels of PAT and SL lipid families were comparatively low in the wild-type strains but markedly higher in their counterparts that carry *M. tuberculosis phoPR*, but the same was not true for ESAT-6, which was secreted at comparable levels in the wild-type and recombinant pairs. The animal-adapted and *M. africanum* L6 strains therefore appear to have acquired compensatory mutations that ameliorate the defect in ESAT-6 production caused by the SNPs in *phoPR*, and so partially restore virulence. In a further twist, convincing evidence was obtained that implicates the insertion of an IS6110 element upstream of *phoPR* in the hypervirulent phenotype of *M. bovis* B - an MDR isolate of *M. bovis* responsible

for an outbreak of TB in Spain [26] - resulting from suppression of the functional deficiencies of the *M. bovis phoPR* allele.

Importantly, this study reinforces the need to sequence additional panels of clinical *M. tuberculosis* isolates as well as other MTBC strains [18] to identify evidence of convergent evolution of functions that might impact bacillary pathogenesis. In contrast to the *M. canettii* group, whose larger genomes have been shaped by extensive inter-strain recombination and horizontal transfer [9,18], the population structure of the MTBC is clonal. It is likely that this clonal restriction, which is evident in the identification of 2,400 SNPs (at most) in a 4.4 Mb MTBC genome, reflects the combined selective pressure of obligate pathogenesis, as well as the close association of MTBC with their natural hosts. In addition, the impact on apparent diversity of strain sampling and laboratory propagation remains unclear. For this reason, the recent use of shotgun metagenomics in clinical TB samples [27] is encouraging, as it suggests that 'culture-free' techniques might enable key insights into the mycobacterial population structure in specific anatomical compartments, while avoiding the biases inherent in existing sample-collection techniques.

### Understanding the genomics of TB drug resistance

In no other area of TB research has the impact of genomics been more profound than in establishing the mechanisms that enable the resistance of *M. tuberculosis* to TB drugs. Like analogous research on other bacterial pathogens, elucidation of the genetic basis of resistance of *M. tuberculosis* to the first-line drugs for the treatment of TB - isoniazid, rifampicin, ethambutol and pyrazinamide - pre-dated the introduction of routine whole-genome sequencing (WGS) of resistant mutants [28]. The discovery that the majority of rifampicin resistance-conferring mutations found in clinical isolates map to an 81-bp region within the *rpoB* gene enabled the development and implementation of the new molecular diagnostic, Xpert MTB/RIF. This test allows for rapid identification of *M. tuberculosis* within clinical specimens and simultaneous identification of rifampicin resistance - a key genetic marker of MDR-TB [29,30]. Assessing the medical, public health and economic impacts of this potentially 'game-changing' technology [31] is the subject of intense investigation in South Africa, where an ambitious program to roll-out Xpert MTB/RIF nationally is underway [32].

More recently, WGS has been used to analyze strains of *M. tuberculosis* with varying drug susceptibility profiles from collections of clinical isolates, as well as drug-resistant mutants isolated in the laboratory [33]. In addition to identifying both canonical resistance-conferring mutations and compensatory mutations, the comparative genomic

analyses of Farhat *et al.* [34] and Zhang *et al.* [35] identified a significant number of new resistance-associated mutations not previously implicated in genetic drug resistance [36]. Their observations suggest that the development of drug resistance in *M. tuberculosis* is a more complex biological phenomenon than previously thought - a notion consistent with emerging trends in other areas of antimicrobial drug resistance [37]. However, the impact of these potentially novel resistance-associated mutations on mycobacterial pathogenesis, and their functional contribution to TB drug resistance, is poorly understood. Validation of the association between genotype and phenotype requires transfer of the resistance-associated mutations into a defined genetic background by means of allelic exchange, a laborious and time-consuming exercise not routinely applied in the TB field, even in the case of resistance-conferring mutations [38]. Therefore, although new techniques such as recombineering offer promise of improved throughput for targeted allelic mutagenesis [39], the genetic validation of resistance-linked mutations is likely to remain a significant challenge. Attempts to confirm inferred associations between specific mutations and observed decreases in drug susceptibility are further complicated by the increasing awareness of the significant capacity of mycobacterial populations for phenotypic heterogeneity in the response to applied drugs [40]. In addition, recent evidence of strain-specific transcriptional phenotypes suggests that genetic background might be of crucial importance in determining the functional consequences of specific mutations [41].

In an impressive illustration of the application of WGS in analyzing the genomics of TB drug resistance, Casali *et al.* [42] investigated the mechanisms underlying the evolution and transmission of TB drug resistance in Russia by sequencing 1,000 *M. tuberculosis* isolates collected prospectively from clinical TB patients. Notably, the major Beijing lineage clades in this collection of strains were found to contain combinations of resistance and compensatory mutations that conferred TB drug resistance while retaining fitness and transmissibility. Traditionally, public health strategies to counter the threat of drug-resistant TB have focused almost entirely on programmatic issues; however, in highlighting the importance of (micro)biological factors in the persistence and spread of MDR and XDR strains within a population [42], this study added a disturbing new dimension to an already daunting challenge.

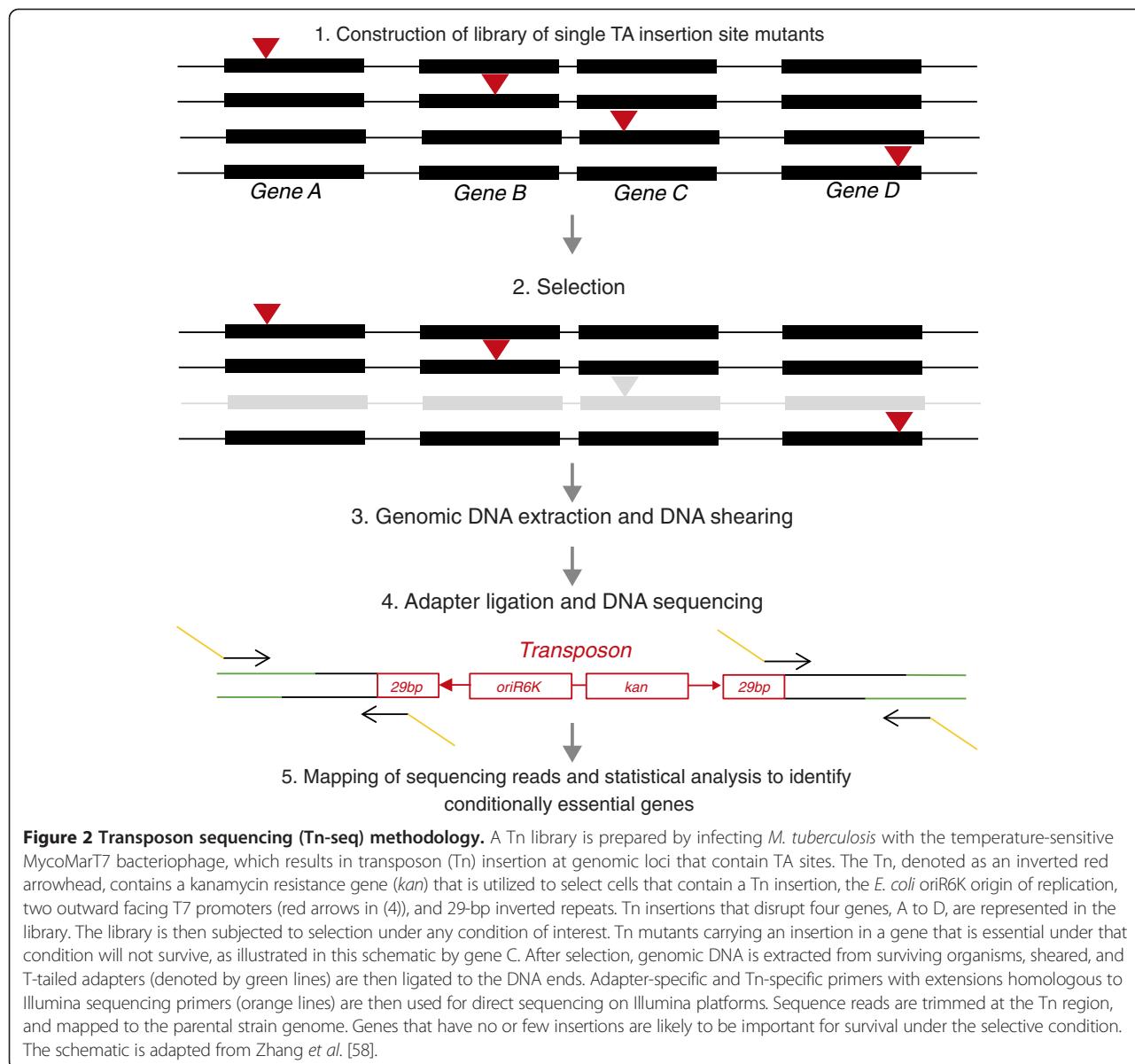
Another area in which WGS analysis of resistant isolates has been applied is in the identification of putative targets of novel anti-mycobacterial agents discovered by screening compound libraries for whole-cell activity against *M. tuberculosis*. This method was successfully used to identify the targets of bedaquiline (the AtpE subunit of ATP synthase), the benzothiazinone BTZ043 (the

DprE1 epimerase), and the imidazopyridine amide Q203 (the QcrB subunit of the respiratory cytochrome *bc*<sub>1</sub> complex) [43]. However, as mutations that compromise drug efficacy frequently map to other resistance-linked genes (such as those encoding efflux pumps) rather than the target, the utility of this method for target identification in *M. tuberculosis* is somewhat limited [39].

### Insights from genome-wide phenotypic profiling of *M. tuberculosis*

As in other fields of microbiology [44], the advent of functional genomics has led to major advances in understanding the biology of *M. tuberculosis* through global phenotypic profiling. This has allowed associations between genotype and phenotype to be uncovered, and has enabled the systematic identification of genes which are required for bacillary growth and survival under conditions that are thought to prevail during human infection. The early application of array-based methods such as transposon site hybridization (TraSH) [45] and signature-tagged mutagenesis [46] provided key insights into the genetic requirements for growth of *M. tuberculosis* *in vitro* [47,48], in macrophages [49], and in animal tissue [50-54]. Recently, these methods have been superseded by transposon sequencing (Tn-Seq), an example of the numerous 'multiletter acronym' or 'MLA-seq' applications [55] that have transformed post-genomic research. In the context of TB, these applications have enabled global phenotypic profiling at significantly higher resolution (Figure 2). Tn-Seq has been used to refine the list of genes required for the growth of *M. tuberculosis* under standard *in vitro* conditions, and to identify the genes needed for growth on cholesterol, a critical carbon source during infection [56]. In an exciting new study that elegantly illustrates the power of this approach, Zhang *et al.* [57] used Tn-Seq to identify sets of genes which the tubercle bacillus engages in order to survive host immunity - so-called 'counteractomes' - thereby uncovering a key role for *de novo* tryptophan biosynthesis in preventing the killing of *M. tuberculosis* by CD4 T cells.

Global phenotypic profiling has been powerfully reinforced by genome-wide transcriptional profiling of *M. tuberculosis* in various experimental models [59-61] and from clinical samples [62,63]. Initially, most analyses utilized DNA microarrays, but RNA sequencing (RNA-Seq) has now largely been adopted as the preferred technique [64-66]. A complex picture is emerging of the manner in which the bacillus modulates its transcriptome in response to environmental cues such as the stresses encountered in the phagosome [67] and the metabolic disturbances caused by chemical inhibition of cellular metabolism [68]. At a practical level, transcriptional profiling has provided a useful tool for categorizing the mechanisms of action of novel anti-mycobacterial agents



[68,69]. From the broader drug discovery perspective, however, the results are ominous: the metabolic flexibility suggested by the various genome-wide transcriptional profiling studies indicates that *M. tuberculosis* is likely to be a difficult target for novel chemotherapies [6]. RNA-Seq has simultaneously uncovered an abundance of non-coding RNAs (ncRNAs) whose expression depends on both physiological stimuli and strain genetic background [41,64]. It seems likely, therefore, that these ncRNAs play a crucial role in the biology of TB infection, as suggested by recent work implicating the PhoP-dependent ncRNA, Mcr7, in regulating the secretion of a key mycobacterial antigen [70].

Comparing genome-wide essentiality and transcriptomic datasets has produced some surprising results: for

example, very little overlap was found between the genes required for survival of *M. tuberculosis* in primary macrophages and those regulated by the intracellular environment, suggesting that gene expression screens may have limited value in identifying virulence genes in pathogens such as *M. tuberculosis* [49]. Nevertheless, the application of these and other genome-scale tools (for example, chromatin immunoprecipitation sequencing (ChIP-Seq) [70,71]), and their integration into systems biology approaches [59], promises to enable a systems-level understanding of the biology of *M. tuberculosis* as an exquisitely adapted human pathogen (Box 1). Importantly, concurrent advances in mycobacterial genetics, chemical biology, cell biology, and imaging have created a powerful platform for the development of

### Box 1. Towards systems biology for tuberculosis

#### A definition of systems biology

The term 'systems biology' is generally used to describe the interacting components of a biological system. Through iterative testing and validation, a mathematical model of the system is constructed, modified, and re-constructed using experimental data obtained from diverse sources. These sources are primarily 'omics applications such as genomics, transcriptomics, proteomics and metabolomics, but also include 'classic' approaches such as molecular biology, genetics, and microbiology. Critically, the model must be able to predict the emergent properties of the system, as well as the impact on the system of external factors and stimuli that might alter specific components or groups of components.

#### Systems biology of TB

The lifecycle of *M. tuberculosis* is driven entirely within the context of human infection: transmission from an infected individual, infection of a new recipient, development of active disease or establishment of a clinically latent state that is able to reactivate, and transmission to a new host. As a result, TB as a disease within an individual might be considered an emergent property of multiple interactions that occur over a range of timescales and at different levels - anatomical, cellular, and molecular - all of which involve elements derived from both bacillus and host. At the level of host populations, systems epidemiology seeks to elucidate the factors - demographic, social, and systemic - that enable the propagation of select *M. tuberculosis* lineages and mutants that are able to survive in the face of control programs and in competition with other genotypes.

#### Some approaches that might be adopted

Direct investigations of mycobacterial function in the context of the complete biological system - the *M. tuberculosis*-infected host - are rare, but will be crucial if the barriers to translating the observations from basic research into practical outcomes are to be overcome. A suite of 'omics techniques must be applied to clinical samples to capture the full diversity of metabolic, proteomic, transcriptomic, and genomic features that characterize the diversity of potentially heterogeneous mycobacterial populations within discrete host compartments and anatomical loci. For example:

- Comparative genomics could be used to identify evidence of convergent evolution in clinical *M. tuberculosis* isolates - both independent of, and associated with, drug resistance.
- Combining and comparing genotypic, epigenetic, and phenotypic data from bacilli captured at different stages of infection - for example, aerosol-encapsulated organisms released by individuals who have active TB versus sputum-based organisms induced for standard clinical diagnostics, paucibacillary populations in immunologically inactive lesions versus bacilli obtained from TB pneumonia, and so on. In all cases, these data should be overlayed with the diversity of host cellular and immunopathological phenotypes.
- Corresponding data should be obtained from experimental models in order to identify the disease-relevant phenotypes and functional interactions that each model system is best able to recapitulate.

novel anti-mycobacterial agents, as well as of diagnostics and biomarkers.

#### Host responses to and biomarkers of *M. tuberculosis* infection

In a parallel approach, post-genomic tools have also been applied in analyzing the response of the human host to infection with *M. tuberculosis*. In a landmark study published in 2010, Ann O'Garra and colleagues [72] identified a 393-gene transcriptional signature in peripheral blood that was able to discriminate patients with active TB from the majority of latently infected and

healthy controls. The neutrophil-driven interferon signature correlated with the extent of disease in those with active TB, as determined by chest X-ray [72], and diminished significantly after only two weeks on standard anti-tubercular therapy, reverting towards that observed in healthy controls [73]. The key findings of this study have been independently validated in different clinical settings and in diverse geographic locations [74-76]. More recently, gene expression signatures have also been identified that distinguish TB from other diseases prevalent in HIV-infected adults [77,78] and in children [79,80]. Together, these observations underscore the potential

utility of blood transcriptional signatures as biomarkers for application in TB diagnosis and in monitoring of response to therapy.

Genomics research also promises to enable significant advances in the discovery of biomarkers and the development of point-of-care diagnostics. The elucidation of a blood transcriptional signature that can identify active TB cases [72] offers the possibility of significantly reducing the diagnostic delay that has been implicated in increased *M. tuberculosis* transmission and the emergence of drug resistance [81]. As noted elsewhere [82], the distinction between active TB and subclinical infection in this transcriptional assay is not absolute, which suggests that this test might be usefully applied to determine the extent of pathology (or bacterial burden) in latently infected individuals, and so might enable identification of those individuals most likely to progress to active disease. To our knowledge, the strength of the transcriptional signature has not been correlated with disease (or bacterial burden). It seems, therefore, that applying an equivalent assay in a non-human primate model [83] might enable calibration of the transcriptional signature against bacillary load and disease pathology. Whether a transcription-based assay of this nature can be applied in resource-limited, disease-endemic regions is currently uncertain; nevertheless, the diagnostic resolution enabled by such approaches suggests that further development is warranted. An additional consequence of these and other transcriptional analyses of host responses to *M. tuberculosis* infection is that fundamental questions have been raised about type I interferon signaling and its role in influencing the outcome of TB infection. As a result, the foundation has been established for systems immunology [84] approaches to understanding the immunopathogenesis of TB, and to developing vaccines and biomarkers through integration with mechanistic studies in cell-based and animal models of infection [85-87].

### **Understanding the genotypic diversity of *M. tuberculosis* within and between hosts**

Advances in high-throughput DNA-sequencing technology have transformed modern bacteriology [88], and their impact on TB genomics has been equally profound [89]. WGS of clinical *M. tuberculosis* isolates has enabled high-resolution insight into strain diversity [6,10], lineage-specific adaptation to host populations [11,12], and micro-variation within hosts and communities [13-15]. In addition to providing strong evidence that bacillary genetics - and, therefore, function - are a significant element in determining the heterogeneous outcomes of infection, these observations suggest that WGS might be profitably incorporated into field trials of new-generation TB interventions, including drugs and vaccines. In one example, a retrospective

observational study [90] that assessed patients from the REMoxTB trial of moxifloxacin-containing drug regimens [91] demonstrated the superiority of WGS over traditional genotyping methods for differentiating cases of relapse and re-infection. This study also confirmed a role for WGS in defining endpoints of clinical trials conducted in high-burden settings. In another example, recent work investigating the intra-patient evolution of *M. tuberculosis* in MDR patients undergoing longitudinal treatment demonstrated long-term co-existence of different bacillary sub-populations [92]. Notably, this study also documented the presence in individual patients of clonal sub-populations that possess different combinations of drug-resistance alleles, a result that has profound implications for phenotypic and molecular drug-resistance testing algorithms, which have traditionally assumed a monomorphic infecting *M. tuberculosis* population.

The growing evidence for genotypic diversity in *M. tuberculosis* impacts epidemiological analyses of strain prevalence and transmission, too. For example, a recent study has shown that the extent of genotypic diversity characterizing bacilli isolated from a single patient can be as great as that observed between samples obtained from patients along a transmission chain [93]. Consistent with earlier evidence from resected lungs [94] and sputum samples [95], the paper by Perez-Lago *et al.* [93] detected intra-patient diversity at both extrapulmonary and respiratory sites, which was interpreted as evidence that variability can be transmitted. As the authors suggest, this result raises important questions about the threshold that should be applied to differentiate relatedness among *M. tuberculosis* isolates for epidemiological analyses, and so renders the inference of transmission events inherently problematic.

To some extent, this difficulty is alleviated in low-incidence settings, especially where bacterial samples are accompanied by thorough clinical and epidemiological metadata. As an example, a retrospective observational study used WGS of archived samples to infer transmission directionality in household outbreaks of TB in the UK Midlands [96]. Again, the authors identified both intra-patient and between-host strain diversity, but the degree of variation was sufficiently limited to enable a framework to be established for the use of WGS data in field epidemiology. Importantly, these results suggested the possible use of WGS data to inform contact tracing, as well as to identify potential 'super-spreaders' - that is, *M. tuberculosis*-infected individuals who might be responsible for a disproportionate number of secondary cases. Even though high-burden settings are likely to pose a special challenge to the application of genomic epidemiology, there is evidence to support the potential of high-resolution genotyping in defining transmission chains independent of drug resistance [97]. This study

from China appears to be the only one of its kind to date in a TB endemic region, but it does suggest the utility of genomic epidemiology, especially where augmented by good clinical, demographic, and social data [98].

### Challenges and perspectives

As an obligate pathogen, *M. tuberculosis* is distinguished from many other infectious organisms (bacterial, viral, and parasitic) which have recourse to non-human reservoirs. Nevertheless, the application of modern genomics techniques in these diverse systems reinforces the potential to elucidate functions and properties that are essential to pathogenesis [99], or which drive the rapid emergence of outbreak strains [100] and ensure their long-term circulation within host populations [101]. High-resolution genotyping, in particular, has revealed that the diversification of clonal infecting strains into 'clouds of diversity' [88] is a feature of many different pathogenic organisms. Determining the extent to which intraspecific diversity is crucial for pathogenesis therefore represents a key research question, and will require the development of systems biology approaches to determine the emergent properties of microdiverse infecting populations.

For TB, it will be useful to consider the immediate research priorities in the context of the major lifecycle stages - active disease, clinical latency, and transmission - and to prioritize genomics applications that are most likely to inform future drug and vaccine development programs (Box 2). The application of advanced 'omic tools is key to novel approaches such as systems epidemiology [102] that aim to combine high-resolution epidemiological data with

systems biology. Nevertheless these techniques must also be harnessed in developing methods for predictive epidemiology that can enable genuinely transformative interventions in TB incidence. As outlined above, the use of WGS to enable definitive differentiation of relapse from re-infection has very significant implications for trials of experimental drug regimens [90]. This is a particularly important consideration in high-burden settings where the force of infection is elevated [2], mixed infections common [103], and a large percentage of recurrent TB is due to exogenous re-infection [104]. Moreover, the potential for epigenetic modifications, such as DNA methylation, to alter bacillary physiology [105] suggests that novel sampling methods and sequencing technologies [100] will be useful in determining the spectrum of physiological states adopted by *M. tuberculosis* within the host and which might impact drug efficacy. Similarly, establishing whether prior infection with one bacillary genotype might predispose to re-infection with a separate genotype following chemotherapeutic elimination [90] is essential, not only for control programs but also for TB vaccine development strategies.

In summary, genomics research will continue to drive efforts to understand the evolutionary processes that have enabled the adaptation of *M. tuberculosis* as a human pathogen. Translating the exciting advances provided by genomics into new tools that can radically transform TB control will require significant and sustained resourcing. It is incumbent upon the TB research community to ensure that there is sufficient political will to make this happen.

### Box 2. Translational priorities

#### *Identifying and intervening in transmission chains*

Can we develop WGS-based methods to identify transmission 'hotspots' and transmission chains to enable real-time interventions to limit the spread of virulent and/or drug-resistant strains?

#### *Identifying the factors that impact infection outcomes*

Can we apply systems biology methods to determine the bacillary and host genetic factors that drive disease progression in specific individuals?

#### *Drug treatment*

Can we utilize WGS-based methods to identify mixed *M. tuberculosis* infections prior to initiation of treatment?

Can we exploit host transcriptional profiling to determine the response to treatment?

#### *Latent infection and vaccinology*

Can we use host transcriptional profiling to identify (and treat) latently infected individuals with a high probability of progressing to active disease?

Can knowledge about mycobacterial diversity be used to guide vaccine development and use in TB-endemic regions?

#### *Mycobacterial population biology and genomics*

Can we determine the impact of intraspecific diversity on disease progression and the emergence of drug resistance?

## Abbreviations

ChIP-Seq: Chromatin immunoprecipitation sequencing; HGT: Horizontal gene transfer; MDR: Multidrug resistant; MTBC: *Mycobacterium tuberculosis* complex; ncRNA: Non-coding RNA; PAT: Polyacyltrehalose; RNA-Seq: RNA sequencing; SL: Sulfolipid; SNP: Single nucleotide polymorphism; TB: Tuberculosis; Tn: Transposon; Tn-Seq: Transposon sequencing; TraSH: Transposon site hybridization; WGS: Whole-genome sequencing; XDR: Extensively drug resistant.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements

We apologize to those authors whose work was not cited because of space limitations. We thank Anastasia Koch for technical assistance and helpful discussions. We acknowledge funding from the National Research Foundation of South Africa and the Howard Hughes Medical Institute (Senior International Research Scholar's grant to VM). Work in our laboratory on TB transmission is funded by the South African Medical Research Council (MRC) with funds from National Treasury under the Economic Competitiveness and Support Package (MRC-RFA-UFSP-01-2013/CCAMP).

## Author details

<sup>1</sup>MRC/NHLS/UCT Molecular Mycobacteriology Research Unit, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, Cape Town 7925, South Africa. <sup>2</sup>DST/NRF Centre of Excellence for Biomedical TB Research, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Anzio Road, Observatory, Cape Town 7925, South Africa. <sup>3</sup>Division of Medical Microbiology, Department of Clinical Laboratory Sciences, University of Cape Town, Anzio Road, Observatory, Cape Town 7925, South Africa.

Published online: 07 November 2014

## References

- Zumla A, George A, Sharma V, Herbert N, Baroness Masham of Ilton: WHO's 2013 global report on tuberculosis: successes, threats, and opportunities. *Lancet* 2013, **382**:1765–1767.
- Wood R, Lawn SD, Johnstone-Robertson S, Bekker LG: Tuberculosis control has failed in South Africa-time to reappraise strategy. *S Afr Med J* 2011, **101**:111–114.
- Dheda K, Gumbo T, Gandhi NR, Murray M, Theron G, Udwadia Z, Migliori GB, Warren R: Global control of tuberculosis: from extensively drug-resistant to untreatable tuberculosis. *Lancet Respir Med* 2014, **2**:321–338.
- Abubakar I, Zignol M, Falzon D, Raviglione M, Ditiu L, Masham S, Adetifa I, Ford N, Cox H, Lawn SD, Marais BJ, McHugh TD, Mwaba P, Bates M, Lipman M, Zijenah L, Logan S, McNerney R, Zumla A, Sarda K, Nahid P, Hoelscher M, Plestschette M, Memish ZA, Kim P, Hafner R, Cole S, Migliori GB, Maeurer M, Schito M, et al: Drug-resistant tuberculosis: time for visionary political leadership. *Lancet Infect Dis* 2013, **13**:529–539.
- Anonymous: *Global Tuberculosis Report 2012*. Geneva: World Health Organization; 2012.
- Warner DF: *Mycobacterium tuberculosis* metabolism. *Cold Spring Harbor Perspect Med* 2014, doi:10.1101/cshperspect.a021121.
- Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaiia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, et al: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998, **393**:537–544.
- Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S: Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 2013, **45**:1176–1182.
- Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, Fiette L, Orgeur M, Fabre M, Parmentier C, Frigui W, Simeone R, Boritsch EC, Debré AS, Willery E, Walker D, Quail MA, Ma L, Bouchier C, Salvignol G, Sayes F, Cascioferro A, Seemann T, Barbe V, Locht C, Gutierrez MC: Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. *Nat Genet* 2013, **45**:172–179.
- Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM, van Pittius NC G: Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerg Infect Dis* 2010, **16**:1296–1299.
- Coscolla M, Lewin A, Metzger S, Maetz-Rensing K, Calvignac-Spencer S, Nitsche A, Dabrowski PW, Radonic A, Niemann S, Parkhill J, Couacy-Hymann E, Feldman J, Comas I, Boesch C, Gagneux S, Leendertz FH: Novel *Mycobacterium tuberculosis* complex isolate from a wild chimpanzee. *Emerg Infect Dis* 2013, **19**:969–976.
- Bos KL, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ, Campbell TJ, Majander K, Wilbur AK, Guichon RA, Wolfe Steadman DL, Cook DC, Niemann S, Behr MA, Zumarraga M, Bastida R, Huson D, Nieselt K, Young D, Parkhill J, Buikstra JE, Gagneux S, Stone AC, Krause J: Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 2014, **514**:494–497.
- Pepperell CS, Casto AM, Kitchen A, Granka JM, Cornejo OE, Holmes EC, Birren B, Galagan J, Feldman MW: The role of selection in shaping diversity of natural *M. tuberculosis* populations. *PLoS Pathog* 2013, **9**:e1003543.
- Galagan JE: Genomic insights into tuberculosis. *Nat Rev Genet* 2014, **15**:307–320.
- Comas I, Chakravarti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD, Gagneux S: Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet* 2010, **42**:498–503.
- Homolka S, Niemann S, Russell DG, Rohde KH: Functional genetic diversity among *Mycobacterium tuberculosis* complex clinical isolates: delineation of conserved core and lineage-specific transcriptomes during intracellular survival. *PLoS Pathog* 2010, **6**:e1000988.
- Fenner L, Egger M, Bodmer T, Furrer H, Ballif M, Battegay M, Helbling P, Fehr J, Gsporer T, Rieder HL, Zwahlen M, Hoffmann M, Bernasconi E, Cavassini M, Calmy A, Dolina M, Frei R, Janssens JP, Borrell S, Stucki D, Schrenzel J, Böttger EC, Gagneux S, Swiss HIV Cohort and Molecular Epidemiology of Tuberculosis Study Groups: HIV infection disrupts the sympathetic host-pathogen relationship in human tuberculosis. *PLoS Genet* 2013, **9**:e1003318.
- Boritsch EC, Supply P, Honore N, Seeman T, Stinear TP, Brosch R: A glimpse into the past and predictions for the future: the molecular evolution of the tuberculosis agent. *Mol Microbiol* 2014, **93**:835–852.
- Becq J, Gutierrez MC, Rosas-Magallanes V, Rauzier J, Gicquel B, Neyrolles O, Deschavanne P: Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. *Mol Biol Evol* 2007, **24**:1861–1871.
- Namouchi A, Didelot X, Schock U, Gicquel B, Rocha EP: After the bottleneck: genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection. *Genome Res* 2012, **22**:721–734.
- Casali N, Nikolayevskyy V, Balabanova Y, Ignatyeva O, Kontsevaya I, Harris SR, Bentley SD, Parkhill J, Nejentsev S, Hoffner SE, Horstmann RD, Brown T, Drobniewski F: Microevolution of extensively drug-resistant tuberculosis in Russia. *Genome Res* 2012, **22**:735–745.
- Behr MA, Waters WR: Is tuberculosis a lymphatic disease with a pulmonary portal? *Lancet Infect Dis* 2014, **14**:250–255.
- Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S: High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol* 2008, **6**:e311.
- Coll F, McNerney R, Guerra-Assuncao JA, Glynn JR, Perdigao J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG: A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun* 2014, **5**:4812.
- Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemard C, Moreau F, Laval F, Daffe M, Martin C, Brosch R, Guilhot C: Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. *Proc Natl Acad Sci U S A* 2014, **111**:11491–11496.
- Rivero A, Marquez M, Santos J, Pinedo A, Sanchez MA, Esteve A, Samper S, Martin C: High rate of tuberculosis reinfection during a nosocomial outbreak of multidrug-resistant tuberculosis caused by *Mycobacterium bovis* strain B. *Clin Infect Dis* 2001, **32**:159–161.
- Doughty EL, Sergeant MJ, Adetifa I, Antonio M, Pallen MJ: Culture-independent detection and characterisation of *Mycobacterium tuberculosis* and M.

- africanum* in sputum samples using shotgun metagenomics on a benchtop sequencer. *PeerJ* 2014, **2**:e585.
- 28. Almeida Da Silva PE, Palomino JC: Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs. *J Antimicrob Chemother* 2011, **66**:1417–1430.
  - 29. Lawn SD, Mwaba P, Bates M, Piatek A, Alexander H, Marais BJ, Cuevas LE, McHugh TD, Zijenah L, Kapata N, Abubakar I, McNeerney R, Hoelscher M, Memish ZA, Migliori GB, Kim P, Maeurer M, Schito M, Zumla A: Advances in tuberculosis diagnostics: the Xpert MTB/RIF assay and future prospects for a point-of-care test. *Lancet Infect Dis* 2013, **13**:349–361.
  - 30. Boehme CC, Nabeta P, Hillemann D, Nicol MP, Shenai S, Krapp F, Allen J, Tahirlı R, Blakemore R, Rustamjee R, Milovic A, Jones M, O'Brien SM, Persing DH, Ruesch-Gerdés S, Gotuzzo E, Rodrigues C, Alland D, Perkins MD: Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med* 2010, **363**:1005–1015.
  - 31. Small PM, Pai M: Tuberculosis diagnosis - time for a game change. *N Engl J Med* 2010, **363**:1070–1071.
  - 32. Meyer-Rath G, Schnippli K, Long L, MacLeod W, Sanne I, Stevens W, Pillay S, Pillay Y, Rosen S: The impact and cost of scaling up GeneXpert MTB/RIF in South Africa. *PLoS One* 2012, **7**:e36966.
  - 33. Warner DF, Mizrahi V: Complex genetics of drug resistance in *Mycobacterium tuberculosis*. *Nat Genet* 2013, **45**:1107–1108.
  - 34. Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR, Gardy JL, Johnston JC, Rodrigues M, Tang PK, Kato-Maeda M, Borowsky ML, Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin EJ, Lander ES, Sabeti PC, Murray M: Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat Genet* 2013, **45**:1183–1189.
  - 35. Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, Liu Z, Li C, Galwey N, Deng J, Zhou Y, Zhu Y, Gao Y, Wang T, Wang S, Huang Y, Wang M, Zhong Q, Zhou L, Chen T, Zhou J, Yang R, Zhu G, Hang H, Zhang J, Li F, Wan K, Wang J, Zhang XE, Bi L: Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nat Genet* 2013, **45**:1255–1260.
  - 36. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S: Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* 2012, **44**:106–110.
  - 37. Andersson DI, Hughes D: Microbiological effects of sublethal levels of antibiotics. *Nat Rev Microbiol* 2014, **12**:465–478.
  - 38. Nebenzahl-Guimaraes H, Jacobson KR, Farhat MR, Murray MB: Systematic review of allelic exchange experiments aimed at identifying mutations that confer drug resistance in *Mycobacterium tuberculosis*. *J Antimicrob Chemother* 2014, **69**:331–342.
  - 39. Ioerger TR, O'Malley T, Liao R, Guinn KM, Hickey MJ, Mohadeen N, Murphy KC, Boshoff HI, Mizrahi V, Rubin EJ, Sasseti CM, Barry CE 3rd, Sherman DR, Parish T, Sacchettini JC: Identification of new drug targets and resistance mechanisms in *Mycobacterium tuberculosis*. *PLoS One* 2013, **8**:e75245.
  - 40. Kester JC, Fortune SM: Persisters and beyond: mechanisms of phenotypic drug resistance and drug tolerance in bacteria. *Crit Rev Biochem Mol Biol* 2014, **49**:91–101.
  - 41. Rose G, Cortes T, Comas I, Coscolla M, Gagneux S, Young DB: Mapping of genotype-phenotype diversity among clinical isolates of *Mycobacterium tuberculosis* by sequence-based transcriptional profiling. *Genome Biol Evol* 2013, **5**:1849–1862.
  - 42. Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, Corander J, Bryant J, Parkhill J, Nejentsev S, Horstmann RD, Brown T, Drobniowski F: Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat Genet* 2014, **46**:279–286.
  - 43. Zumla A, Nahid P, Cole ST: Advances in the development of new tuberculosis drugs and treatment regimens. *Nat Rev Drug Discov* 2013, **12**:388–404.
  - 44. van Opijken T, Camilli A: Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* 2013, **11**:435–442.
  - 45. Murry JP, Sasseti CM, Lane JM, Xie Z, Rubin EJ: Transposon site hybridization in *Mycobacterium tuberculosis*. *Methods Mol Biol* 2008, **416**:45–59.
  - 46. Cox JS, Chen B, McNeil M, Jacobs WR Jr: Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. *Nature* 1999, **402**:79–83.
  - 47. Sasseti CM, Boyd DH, Rubin EJ: Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 2003, **48**:77–84.
  - 48. Lamichhane G, Zignol M, Blades NJ, Geiman DE, Dougherty A, Grosset J, Broman KW, Bishai WR: A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 2003, **100**:7213–7218.
  - 49. Rengarajan J, Bloom BR, Rubin EJ: Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc Natl Acad Sci U S A* 2005, **102**:8327–8332.
  - 50. Sasseti CM, Rubin EJ: Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* 2003, **100**:12989–12994.
  - 51. Sasseti CM, Boyd DH, Rubin EJ: Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci U S A* 2001, **98**:12712–12717.
  - 52. Lamichhane G, Tyagi S, Bishai WR: Designer arrays for defined mutant analysis to detect genes essential for survival of *Mycobacterium tuberculosis* in mouse lungs. *Infect Immun* 2005, **73**:2533–2540.
  - 53. Jain SK, Hernandez-Abanto SM, Cheng QJ, Singh P, Ly LH, Klinkenberg LG, Morrison NE, Converse PJ, Nuernberger E, Grosset J, McMurray DN, Karakousis PC, Lamichhane G, Bishai WR: Accelerated detection of *Mycobacterium tuberculosis* genes essential for bacterial survival in guinea pigs, compared with mice. *J Infect Dis* 2007, **195**:1634–1642.
  - 54. Dutta NK, Mehra S, Didier PJ, Roy CJ, Doyle LA, Alvarez X, Ratterree M, Be NA, Lamichhane G, Jain SK, Lacey MR, Lackner AA, Kaushal D: Genetic requirements for the survival of tubercle bacilli in primates. *J Infect Dis* 2010, **201**:1743–1752.
  - 55. McPherson JD: A defining decade in DNA sequencing. *Nat Methods* 2014, **11**:1003–1005.
  - 56. Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sasseti CM: High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* 2011, **7**:e100251.
  - 57. Zhang YJ, Reddy MC, Ioerger TR, Rothchild AC, Dartois V, Schuster BM, Trauner A, Wallis D, Galaviz S, Huttenhower C, Sacchettini JC, Behar SM, Rubin EJ: Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. *Cell* 2013, **155**:1296–1308.
  - 58. Zhang YJ, Ioerger TR, Huttenhower C, Long JE, Sasseti CM, Sacchettini JC, Rubin EJ: Global assessment of genomic regions required for growth in *Mycobacterium tuberculosis*. *PLoS Pathog* 2012, **8**:e1002946.
  - 59. Rohde KH, Veiga DF, Caldwell S, Balazsi G, Russell DG: Linking the transcriptional profiles and the physiological states of *Mycobacterium tuberculosis* during an extended intracellular infection. *PLoS Pathog* 2012, **8**:e1002769.
  - 60. Griffin JE, Pandey AK, Gilmore SA, Mizrahi V, McKinney JD, Bertozzi CR, Sasseti CM: Cholesterol catabolism by *Mycobacterium tuberculosis* requires transcriptional and metabolic adaptations. *Chem Biol* 2012, **19**:218–227.
  - 61. Talaat AM, Lyons R, Howard ST, Johnston SA: The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. *Proc Natl Acad Sci U S A* 2004, **101**:4602–4607.
  - 62. Rachman H, Strong M, Ulrichs T, Grode L, Schuchhardt J, Mollenkopf H, Kosmiadi GA, Eisenberg D, Kaufmann SH: Unique transcriptome signature of *Mycobacterium tuberculosis* in pulmonary tuberculosis. *Infect Immun* 2006, **74**:1233–1242.
  - 63. Garton NJ, Waddell SJ, Sherratt AL, Lee SM, Smith RJ, Senner C, Hinds J, Rajakumar K, Adegbola RA, Besra GS, Butcher PD, Barer MR: Cytological and transcript analyses reveal fat and lazy persister-like bacilli in tuberculous sputum. *PLoS Med* 2008, **5**:e75.
  - 64. Arnvig KB, Comas I, Thomson NR, Houghton J, Boshoff HI, Croucher NJ, Rose G, Perkins TT, Parkhill J, Dougan G, Young DB: Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*. *PLoS Pathog* 2011, **7**:e1002342.
  - 65. Cortes T, Schubert OT, Rose G, Arnvig KB, Comas I, Aebersold R, Young DB: Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Rep* 2013, **5**:1121–1131.
  - 66. Upolekar S, Rougemont J, Cole ST, Sala C: High-resolution transcriptome and genome-wide dynamics of RNA polymerase and NusA in *Mycobacterium tuberculosis*. *Nucleic Acids Res* 2013, **41**:961–977.
  - 67. Schnappinger D, Ehrhart S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK: Transcriptional

- adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. *J Exp Med* 2003, **198**:693–704.
- 68. Boshoff HI, Myers TG, Copp BR, McNeil MR, Wilson MA, Barry CE 3rd: The transcriptional responses of *Mycobacterium tuberculosis* to inhibitors of metabolism: novel insights into drug mechanisms of action. *J Biol Chem* 2004, **279**:40174–40184.
  - 69. Murima P, de Sessions PF, Lim V, Naim AN, Bifani P, Boshoff HI, Sambandamurthy VK, Dick T, Hibberd ML, Schreiber M, Rao SP: Exploring the mode of action of bioactive compounds by microfluidic transcriptional profiling in mycobacteria. *PLoS One* 2013, **8**:e69191.
  - 70. Solans L, Gonzalo-Asensio J, Sala C, Benjak A, Uplekar S, Rougemont J, Guilhot C, Malaga W, Martin C, Cole ST: The PhoP-dependent ncRNA Mcr7 modulates the TAT secretion system in *Mycobacterium tuberculosis*. *PLOS Pathog* 2014, **10**:e1004183.
  - 71. Galagan JE, Minch K, Peterson M, Lyubetskaya A, Azizi E, Sweet L, Gomes A, Rustad T, Dolganov G, Glotova I, Abeel T, Mahwinney C, Kennedy AD, Allard R, Brabant W, Krueger A, Jaini S, Honda B, Yu WH, Hickey MJ, Zucker J, Garay C, Weiner B, Sisk P, Stolte C, Winkler JK, Van de Peer Y, Iazzetti P, Camacho D, Dreyfuss J: The *Mycobacterium tuberculosis* regulatory network and hypoxia. *Nature* 2013, **499**:178–183.
  - 72. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, Wilkinson KA, Banchereau R, Skinner J, Wilkinson RJ, Quinn C, Blankenship D, Dhawan R, Cush JJ, Mejias A, Ramilo O, Kon OM, Pascual V, Banchereau J, Chaussabel D, O'Garra A: An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010, **466**:973–977.
  - 73. Bloom CI, Graham CM, Berry MP, Wilkinson KA, Oni T, Rozakeas F, Xu Z, Rossello-Urgell J, Chaussabel D, Banchereau J, Pascual V, Lipman M, Wilkinson RJ, O'Garra A: Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS One* 2012, **7**:e46191.
  - 74. Maertzdorf J, Replibier D, Parida SK, Stanley K, Roberts T, Black G, Walz G, Kaufmann SH: Human gene expression profiles of susceptibility and resistance in tuberculosis. *Genes Immun* 2011, **12**:15–22.
  - 75. Ottenhoff TH, Dass RH, Yang N, Zhang MM, Wong HE, Sahiratmadja E, Khor CC, Alisjahbana B, van Crevel R, Marzuki S, Seielstad M, van de Vosse E, Hibberd ML: Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. *PLoS One* 2012, **7**:e45839.
  - 76. Cliff JM, Lee JS, Constantinou N, Cho JE, Clark TG, Ronacher K, King EC, Lukey PT, Duncan K, Van Helden PD, Walz G, Dockrell HM: Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. *J Infect Dis* 2013, **207**:18–29.
  - 77. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, Banwell CM, Brent AJ, Crampin AC, Dockrell HM, Eley B, Heyderman RS, Hibberd ML, Kern F, Langford PR, Ling L, Mendelson M, Ottenhoff TH, Zgambro F, Wilkinson RJ, Coin LJ, Levin M: Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med* 2013, **10**:e1001538.
  - 78. Dawany N, Showe LC, Kossenkov AV, Chang C, Iive P, Conradie F, Stevens W, Sanne I, Azzioni L, Montaner LJ: Identification of a 251 gene expression signature that can accurately detect *M. tuberculosis* in patients with and without HIV co-infection. *PLoS One* 2014, **9**:e89925.
  - 79. Anderson ST, Kaforou M, Brent AJ, Wright VJ, Banwell CM, Chagaluka G, Crampin AC, Dockrell HM, French N, Hamilton MS, Hibberd ML, Kern F, Langford PR, Ling L, Mlotha R, Ottenhoff TH, Pienaar S, Pillay V, Scott JA, Twahir H, Wilkinson RJ, Coin LJ, Heyderman RS, Levin M, Eley B, ILULU Consortium; KIDS TB Study Group: Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med* 2014, **370**:1712–1723.
  - 80. Cattamanchi A, Walter ND, Metcalfe JZ, Davis JL: A transcriptional signature for active TB: have we found the needle in the haystack? *PLoS Med* 2013, **10**:e1001539.
  - 81. MacPherson P, Houben RM, Glynn JR, Corbett EL, Kranzer K: Pre-treatment loss to follow-up in tuberculosis patients in low- and lower-middle-income countries and high-burden countries: a systematic review and meta-analysis. *Bull World Health Organ* 2014, **92**:126–138.
  - 82. Fortune SM, Rubin EJ: Host transcription in active and latent tuberculosis. *Genome Biol* 2010, **11**:135.
  - 83. Lin PL, Rodgers M, Smith L, Bigbee M, Myers A, Bigbee C, Chiosea I, Capuano SV, Fuhrman C, Klein E, Flynn JL: Quantitative comparison of active and latent tuberculosis in the cynomolgus macaque model. *Infect Immun* 2009, **77**:4631–4642.
  - 84. Rappuoli R, Aderem A: A 2020 vision for vaccines against HIV, tuberculosis and malaria. *Nature* 2011, **473**:463–469.
  - 85. Blankley S, Berry MP, Graham CM, Bloom CI, Lipman M, O'Garra A: The application of transcriptional blood signatures to enhance our understanding of the host response to infection: the example of tuberculosis. *Philos Trans R Soc Lond B Biol Sci* 2014, **369**:20130427.
  - 86. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MP: The immune response in tuberculosis. *Annu Rev Immunol* 2013, **31**:475–527.
  - 87. Andersen P, Woodworth JS: Tuberculosis vaccines-rethinking the current paradigm. *Trends Immunol* 2014, **35**:387–395.
  - 88. McAdam PR, Richardson EJ, Fitzgerald JR: High-throughput sequencing for the study of bacterial pathogen biology. *Curr Opin Microbiol* 2014, **19**:C106–113.
  - 89. Gagneux S: Genetic diversity in *Mycobacterium tuberculosis*. *Curr Top Microbiol Immunol* 2013, **374**:1–25.
  - 90. Bryant JM, Harris SR, Parkhill J, Dawson R, Diacon AH, van Helden P, Pym A, Mahayiddin AA, Chuchottaworn C, Sanne IM, Louw C, Boeree MJ, Hoelscher M, McHugh TD, Bateson AL, Hunt RD, Mwaigwisya S, Wright L, Gillespie SH, Bentley SD: Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respir Med* 2013, **1**:786–792.
  - 91. Gillespie SH, Crook AM, McHugh TD, Mendel CM, Meredith SK, Murray SR, Pappas F, Phillips PP, Nunn AJ: Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis. *N Engl J Med* 2014, **371**:1577–1587.
  - 92. Merker M, Kohl TA, Roetzer A, Trübele B, Richter E, Rusch-Gerdes S, Fattorini L, Oggioni MR, Cox H, Varaine F, Niemann S: Whole genome sequencing reveals complex evolution patterns of multidrug-resistant *Mycobacterium tuberculosis* Beijing strains in patients. *PLoS One* 2013, **8**:e82551.
  - 93. Perez-Lago L, Comas I, Navarro Y, Gonzalez-Candelas F, Herranz M, Bouza E, Garcia-de-Viedma D: Whole genome sequencing analysis of intrapatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. *J Infect Dis* 2014, **209**:98–108.
  - 94. Kaplan G, Post FA, Moreira AL, Wainwright H, Kreiswirth BN, Tanverdi M, Mathema B, Ramaswamy SV, Walther G, Steyn LM, Barry CE 3rd, Bekker LG: *Mycobacterium tuberculosis* growth at the cavity surface: a microenvironment with failed immunity. *Infect Immun* 2003, **71**:7099–7108.
  - 95. Sun G, Luo T, Yang C, Dong X, Li J, Zhu Y, Zheng H, Tian W, Wang S, Barry CE 3rd, Mei J, Gao Q: Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J Infect Dis* 2012, **206**:1724–1733.
  - 96. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE: Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 2013, **13**:137–146.
  - 97. Luo T, Yang C, Peng Y, Lu L, Sun G, Wu J, Jin X, Hong J, Li F, Mei J, DeRiemer K, Gao Q: Whole genome sequencing to detect recent transmission of *Mycobacterium tuberculosis* in settings with a high burden of tuberculosis. *Tuberculosis* 2014, **94**:434–440.
  - 98. Middelkoop K, Mathema B, Myer L, Shashkina E, Whitelaw A, Kaplan G, Kreiswirth B, Wood R, Bekker LG: Transmission of tuberculosis in a South African community with a high prevalence of HIV infection. *J Infect Dis* 2014.
  - 99. Kafsack BF, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, Williams AE, Drought LG, Kwiatkowski DP, Baker DA, Cortés A, Llinás M: A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* 2014, **507**:248–252.
  - 100. Fang G, Munera D, Friedman DL, Mandlik A, Chao MC, Banerjee O, Feng Z, Losic B, Mahajan MC, Jabado OJ, Deikus G, Clark TA, Luong K, Murray IA, Davis BM, Keren-Paz A, Chess A, Roberts RJ, Korlach J, Turner SW, Kumar V, Waldor MK, Schadt EE: Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat Biotechnol* 2012, **30**:1232–1239.
  - 101. Gire SK, Goba A, Andersen KG, Sealton RSG, Park DJ, Kanneh L, Jaloh S, Momoh M, Fullah M, Dudas G, Wohl S, Moses LM, Yozwiak NL, Winnicki S, Matranga CB, Malboeuf CM, Qu J, Gladden AD, Schaffner SF, Yang X, Jiang PP, Nekoui M, Colubri A, Coomber MR, Fonnlie M, Moigboi A, Gbakie M, Kamara FK, Tucker V, Konuwa E, et al: Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 2014, **345**:1369–1372.
  - 102. Comas I, Gagneux S: A role for systems epidemiology in tuberculosis research. *Trends Microbiol* 2011, **19**:492–500.
  - 103. Hanekom M, Streicher EM, Van de Berg D, Cox H, McDermid C, Bosman M, van Pittius NC G, Victor TC, Kidd M, van Soolingen D, van Helden PD,

- Warren RM: Population structure of mixed *Mycobacterium tuberculosis* infection is strain genotype and culture medium dependent. *PLoS One* 2013, 8:e70178.
104. Verver S, Warren RM, Beyers N, Richardson M, van der Spuy GD, Borgdorff MW, Enarson DA, Behr MA, van Helden PD: Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. *Am J Respir Crit Care Med* 2005, 171:1430–1435.
105. Shell SS, Prestwich EG, Baek SH, Shah RR, Sassetti CM, Dedon PC, Fortune SM: DNA methylation impacts gene expression and ensures hypoxic survival of *Mycobacterium tuberculosis*. *PLoS Pathog* 2013, 9:e1003419.

doi:10.1186/s13059-014-0514-z

**Cite this article as:** Warner and Mizrahi: Translating genomics research into control of tuberculosis: lessons learned and future prospects. *Genome Biology* 2014 15:514.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

