Released Online: August 21, 2016

COMPASS-TB Report Design Study: First Online Survey

Geoff McKee, Ana Crisan, Jennifer Gardy, Tamara Munzner

Summary

- Most respondents (15/17) from UK
- Majority (10/17) involved in clinical management, 8/17 involved in epidemiology/surveillance
- Majority of respondents felt at least somewhat confident interpreting genome-derived data types
- Most respondents involved in clinical management (5/6) wanted data quality metrics
- Most commonly required genome-derived data types for clinical management included speciation, predicted drug sensitivity, and resistotype
- Less than 50% reported requiring genome-derived data for each of the epidemiological tasks, aside from the use of speciation for contact tracing

Introduction:

Many public health agencies are starting to use whole genome sequencing as a tool for diagnosing infections, predicting drug resistance, and identifying closely related isolates that might suggest an outbreak. The COMPASS-TB study compared a novel WGS diagnostic pipeline to the standard diagnostic approach and demonstrated that, when this technique is used in the tuberculosis laboratory, we can generate all the usual results that one has come to expect from a reference mycobacteriology lab, but we can do so much faster and at lower cost (Pankhurst *et al.*, 2016). As a result of this study, groups like Public Health England, the BC Centre for Disease Control, and the US Centers for Disease Control and Prevention are all using genomics to analyze their incoming mycobacterial isolates.

The COMPASS-TB study demonstrated that WGS can be used to diagnose, resistance type and genotype tuberculosis isolates more efficiently than the current diagnostic pipeline; however, further work was required to translate these often complex results into something that can be interpreted and used by clinicians and public health professionals. This study aims to use formal InfoVis methodologies to develop a report for the COMPASS-TB diagnostic WGS pipeline (Sedlmair *et al.*, 2012). The design of the final report will be guided by user preference, while meeting both their needs and reporting requirements for accreditation (International Organization for Standardization, 2012).

In the first phase of the report design study, users of lab data were surveyed to assess knowledge of different data types and how they were used in routine clinical and public health management of tuberculosis.

Methodology:

Participants were recruited from Canada and the UK through convenience and snowball sampling. As an incentive, participants were entered into a draw for a $150 Apple Store gift card. Survey development was guided by the results of expert consultation, as well as tasks derived from Canadian and UK national guidelines for tuberculosis management (Public Health Agency of Canada, 2013; NICE, 2016). The survey focused on participant demographics, understanding of genomic tests for tuberculosis, use of data types for clinical and epidemiological tasks and characteristics of the current tuberculosis lab reporting system. The survey involved 30 questions that were predominantly quantitative, comprised of multiple choice, multiple answer and free text formats. Surveys were piloted among members of the research team for comprehension and then distributed online via fluidsurveys.

Data was collected between July 13th and July 20th, 2016. Results were analyzed using descriptive statistical methods in excel and the fluidsurveys analytic platform.

Ethics approval was obtained through the UBC Human Research Ethics Board (UBC Rise Number: H10-03336).

Results:

A total of 17 complete responses were collected during the survey period. As illustrated in Table 1, the majority of respondents reported involvement in clinical management (10/17), primarily as physicians/clinicians (7/10). Eight respondents reported involvement in surveillance/epidemiology, while others were involved in laboratory work and research. Respondents were predominantly from the United Kingdom and worked in Public Health Organizations and/or Hospitals. While all respondents had heard about or worked on a research project involving WGS, over 50% had no training in bioinformatics; however, the majority of respondents expressed that they were

		# of Respondents (n=17)
Role		
	Clinical Management	10
	Laboratory Work	4
	Surveillance/Epidemiology	8
	Research	4
	Other	2
Country		
	United Kingdom	15
	Canada	1
	United States	0
	Other	1

Table 1. Respondent Demographics

somewhat confident or confident in their ability to interpret genome-derived data types (Figure 1).

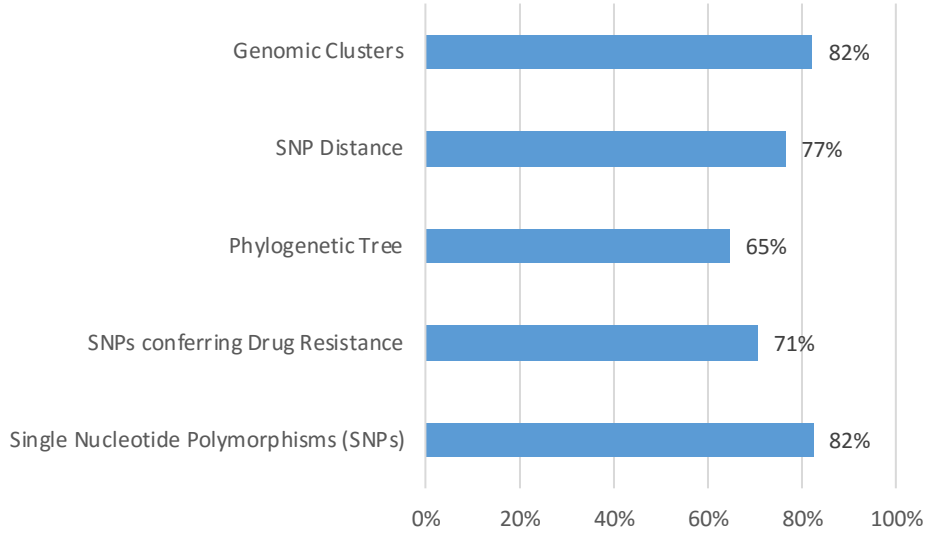

Figure 1. Percent of Respondents somewhat confident or confident in ability to interpret data type (n = 17)

Six respondents were involved in the diagnosis and treatment of tuberculosis and reported receiving lab reports in paper and/or electronic formats. Of these 6 respondents, 5 wanted the data quality metrics associated with lab reports. The requirements for genome-derived data types varied among diagnostic and treatment tasks, with speciation, predicted drug sensitivities, and resistotype among the most commonly requested (Table 2). The main challenges to diagnosis and treatment included timeliness of reports and the disconnect between necessary data reported in different documents.

Twelve respondents reported involvement in the epidemiological aspects of tuberculosis management. Only 7 out of 12 of these respondents reported that they directly review original lab reports. Three of the five who reported not reviewing lab reports received extracted data. Less than 50% reported requiring genome-derived data for all of the epidemiological tasks, aside from the use of speciation for contact tracing (Table 2).

Five respondents reported involvement in tuberculosis surveillance, all of whom also worked for institutions planning to use more genomic data in the future. As demonstrated in Table 2, all 5 respondents described speciation and cluster assignment as required data types for surveillance.

Task	Speciation	Predicted DS	Resistotype	Cluster Assignment	SNP Distance	Phylogenetic Tree	Quality Metrics
Dx Latent TB (n=6)	43%	14%	14%	14%	14%	14%	14%
Dx Active TB (n=6)	71%	43%	43%	43%	29%	43%	29%
Reactivation vs. New Acquisition (n=6)	57%	57%	71%	57%	57%	43%	57%
Transmission Risk (n=6)	71%	43%	43%	43%	29%	43%	29%
Choose Rx (n=6)	86%	86%	86%	29%	29%	29%	29%
Choose Tx Duration (n=6)	71%	71%	86%	29%	29%	29%	29%
Assess Tx Response (n=6)	86%	57%	57%	14%	14%	14%	14%
Contact Tracing (n=12)	53%	29%	35%	41%	35%	24%	24%
Report to PH (n=12)	35%	18%	35%	41%	29%	12%	6%
Define a Cluster (n=12)	35%	12%	35%	41%	41%	24%	18%
Connect Cases to Cluster (n=12)	35%	12%	35%	35%	41%	29%	12%
Guide PH Response (n=12)	41%	18%	35%	41%	35%	29%	12%
Surveillance (n=5)	100%	20%	20%	100%	20%	40%	20%

Table 2. Percent of respondents indicating genome-derived data types required to perform different clinical and epidemiological tasks for management of tuberculosis

Discussion:

The responses for the questions that evaluated the required data types for each task were difficult to interpret as there appeared to be considerable variation. Among the diagnostic tasks, genome-derived data types appeared less useful for diagnosis of latent tuberculosis, although participants were divided on their utility for the other tasks. As hypothesized, speciation appeared useful for treatment tasks, while predicted drug sensitivity and resistotype were also linked with choosing the type and duration of treatment. Respondents appeared to

have less use for cluster or relatedness data types for clinical tasks. This may suggest that clinicians do not routinely have access to clinical data from related cases or they do not find this data useful to guide their clinical management.

Fewer respondents found these data types useful for epidemiological tasks, including the use of cluster assignment and SNP distance for cluster identification. These results were unexpected and may reflect a lack of knowledge about the use of WGS molecular epidemiological techniques, even though this does not correlate with the results of the previous questions evaluating confidence in interpretation of these data types. This could also suggest that genome-derived cluster data may be viewed as supplementary to existing data and not required for epidemiological investigation.

In contrast to the key informant consultations, the majority of respondents involved in clinical management wanted to know about laboratory or bioinformatics quality metrics that may be associated with the reported data. This could reflect a higher level of technical expertise among the survey respondents, although further investigation may be required to determine the need for this data among the general user population.

Interpretation of the survey was limited by the small number of respondents, particularly for the analysis of role-specific tasks. The high proportion of UK respondents also may impede the generalizability of conclusions for other jurisdictions. As demonstrated by the level of enthusiasm indicated by respondents for whole genome sequencing and their ability to interpret genome-derived data types, the conclusions may be influenced by participation bias, with the population representing early-adopters rather than the general population of users.

Next Steps:

The results of this survey will guide the development of prototype report designs by illustrating which data types users require for specific tasks. The prototypes will be developed in collaboration with human computer interaction researchers in a one day “design sprint” workshop. Following this session, the prototypes will be evaluated in a second online survey using discrete choice experimental methods.

References:

International Organization for Standardization. (2012). ISO 15189:2012 - Medical laboratories -- Requirements for quality and competence. Retrieved from http://www.iso.org/iso/catalogue_detail?csnumber=56115

National Institute for Health and Care Excellence. (2016). Tuberculosis: NICE Guideline. Retrieved from https://www.nice.org.uk/guidance/ng33

Pankhurst, L. J., Del Ojo Elias, C., Votintseva, A. A., Walker, T. M., Cole, K., Davies, J., … Crook, D. W. (2016). Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. *The Lancet. Respiratory Medicine*, *4*(1), 49–58. http://doi.org/10.1016/S2213-2600(15)00466-X

Public Health Agency of Canada, The Lung Association, & Canadian Thoracic Association. (2013). Canadian Tuberculosis Standards, 7th Edition. Retrieved from http://www.respiratoryguidelines.ca/sites/all/files/Canadian_TB_Standards_7th_Edition_ENG.pdf

Sedlmair, M., Meyer, M., & Munzner, T. (2012). Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2431–40. http://doi.org/10.1109/TVCG.2012.213