

Appendix D

Genotyping Background Information and Glossary

Tuberculosis (TB) genotyping is a laboratory-based analysis of the genetic material of the bacteria that cause TB disease, *Mycobacterium tuberculosis* complex. The total genetic content is referred to as the genome. Specific sections of the genome contain distinct genetic patterns that help distinguish different strains of *M. tuberculosis*. TB genotyping examines the location, number, and presence of different types of spacer or repetitive DNA patterns. The areas of the genome examined in TB genotyping are different from those related to drug resistance.

Applications of Genotyping

Persons with TB disease who are related by transmission should have matching genotype results. Conversely, persons with matching TB genotyping results are probably related by transmission in some way, although the connection might not be recent or direct.

Genotyping results, when combined with epidemiologic data, can help identify persons with TB disease involved in the same chain of transmission. This information adds value to conventional TB control activities in different ways. These applications are summarized as follows:

Patient-Level Applications of Genotyping

Complete Contact Investigations

- Confirm or refute patient connections (epidemiologic linkages) identified that might or might not be identified through routine contact investigations.

Cluster Investigations

- Find patient connections that were not identified through routine contact investigations.
- Detect, refute, or confirm potential false-positive culture results.
- Distinguish relapse TB disease from new TB infection among TB patients with recurrent TB disease.

Population-Level Applications of Genotyping

- Detect potential outbreaks by using geospatial or other analyses of genotype clusters.
- Refute outbreaks when cases believed to be part of the same outbreak have nonmatching genotype results.
- Define the scope of potential outbreaks by identifying all cases in an area with a matching genotype.
- Monitor known outbreaks over time by watching for new cases with the outbreak genotype that become added to existing clusters (outbreak surveillance).

History of TB Genotyping Surveillance in the United States

In 1996, CDC started the National Tuberculosis Genotyping Surveillance Network (NTGSN), a 5-year initiative that established the utility of genotyping in TB control efforts.¹ In 2004, based on the knowledge gained from NTGSN and associated studies,² CDC established the National TB Genotyping Service (NTGS) and funded a national genotyping laboratory, located in Michigan, to genotype at least one *M. tuberculosis* isolate from each culture-positive TB case reported in the United States.³ All TB control programs can use NTGS at no cost to the patients, health

care providers, or health departments. NTGS participation is voluntary, with individual programs determining how genotyping data will be used for their TB control activities. Since 2004, approximately 115,000 *M. tuberculosis* isolates have been successfully genotyped through NTGS and its partnerships among CDC programs, national genotyping laboratories, and 58 states and jurisdictions.

In 2010, CDC launched the TB Genotyping Information Management System (TB GIMS), a secure Internet-based database available to all 50 states, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, and the U.S.-affiliated Pacific Islands. TB GIMS makes genotyping data easily available to users and links genotyping data to patient surveillance records. Key features include tools to link genotype results of isolate records from NTGS to patient surveillance records from the National TB Surveillance System (NTSS). Additional features include database queries regarding genotypes and clusters, data quality checks, aggregate reports, maps, and outbreak detection tools. TB GIMS has >500 users among local, state, federal, and territorial partners.

Genotyping-Based Outbreak Detection

CDC identifies genotype clusters that are most likely to represent TB outbreaks. Genotyping-based outbreak detection involves using geospatial analysis to identify unusual groupings of TB cases with matching genotypes that might represent outbreaks. TB control programs can use outbreak detection information to help allocate and prioritize resources for investigation and intervention on specific TB genotype clusters.

CDC's primary outbreak detection method is based on identifying higher than expected geospatial concentrations of a TB genotype in a specific county, compared with the national distribution of that genotype. This method calculates a log-likelihood ratio (LLR) statistic; clusters with higher LLRs are more likely to represent greater geospatial concentrations than clusters with lower LLRs; higher LLRs might indicate recent transmission of TB. LLRs are then classified into alert levels within TB GIMS on the basis of established cut points. Clusters are classified as *no alert* (LLRs 0–<5), *medium alert* (LLRs 5–<10), or *high alert* (≥ 10). The alert level and changes in alert levels (e.g., from no to medium or high) can help TB programs identify outbreaks and prioritize TB genotype clusters for further investigation or intervention.

Genotyping Terminology

In NTGS, a genotype is defined as a unique combination of spacer oligonucleotide typing results (spoligotype) and 24-locus mycobacterial interspersed repetitive unit–variable number tandem repeat typing (MIRU–VNTR) results. Each unique combination of results is assigned a GEN-Type designated as *G* followed by 5 digits, which are assigned sequentially to every genotype identified in the United States (e.g., G00162). This nomenclature is designed for convenience and ease of communication, but the specific numbers assigned have no additional importance outside NTGS. Genotyping data from NTGS should not be used for clinical decision making.

National TB Genotyping Surveillance Coverage in the United States

National TB genotyping surveillance coverage refers to the proportion of culture-positive TB cases with a genotyped *M. tuberculosis* isolate. High levels of coverage in the United States can provide a better understanding of the epidemiology of TB transmission within a specific geographic area, as well as nationally. Additionally, because outbreak detection algorithms are based on identifying unusual geospatial concentrations of genotypes, high coverage levels help

decrease the likelihood of false-negative alerts. The National Tuberculosis Indicator Project national genotyping surveillance coverage objective is 94%.⁴

GLOSSARY

Alert level: A mechanism used by TB GIMS to notify users of genotype clusters, possibly representing TB outbreaks, in a specific county. The alert level is determined by the LLR for a given cluster. This is calculated by TB GIMS and is updated whenever a new case is added to a genotype cluster. E-mail notifications are generated whenever an alert level changes from a no alert LLR ($0 < 5.0$) to medium LLR ($5.0 < 10.0$) or high LLR (≥ 10.0), or from a medium LLR to a high LLR.

Cluster investigation: A cluster investigation identifies epidemiologic links between TB patients whose isolates have matching genotypes. It might consist of reviewing information from public health and medical records and interviewing case managers and outreach workers. It can also involve re-interviewing TB patients.

Epidemiologic link (epi link): An epidemiologic link is a relationship that two TB patients share that explains where, when, and how *M. tuberculosis* might have been transmitted between them. Patients who name each other as contacts have an epidemiologic link. However, an epidemiologic link can be a location where the two persons spent time together or an activity occurred that brought them together.

Genotype: The designation that represents one or more of the three genotyping techniques used for *M. tuberculosis*: spoligotyping, MIRU-VNTR analysis, and IS6110-based restriction fragment length polymorphism (RFLP). These designations were developed to facilitate communication of genotyping information within and between TB programs. In the United States, we use GENType or PCRType to define a genotype.

Genotype surveillance coverage: Genotyping surveillance coverage is defined as the proportion of culture-positive TB cases with a genotype result.

GENType: A designation for each unique combination of spoligotype and 24-locus MIRU-VNTR results. GENType is designated as *G* followed by five digits, which are assigned sequentially to every genotype identified in the United States (e.g., G00017).

Genotyping cluster: A genotyping cluster consists of two or more cases in a jurisdiction during a specified period with *M. tuberculosis* isolates that share matching genotypes. In the United States, all cases with matching GENType or PCRType are considered to be in a genotype cluster. The jurisdiction and period used vary on the basis of the specific application of the term *cluster*. Within TB GIMS, a single county and a 3-year period are used to define a cluster.

Geospatial concentration: Geospatial concentration is a measure of how concentrated a genotype is in time and space. It indicates that recent transmission has occurred because patients with infections with the same genotype in the same location are more likely to have come in contact with each other. TB GIMS uses the LLR to generate a numeric measure of geospatial concentration of a given TB genotype.

Linking: In TB GIMS, *linking* refers to the process of connecting genotyping results with a reported TB case from the National TB Surveillance System (NTSS). This step is essential for ensuring that demographic, risk factor, and geographic data can be viewed in TB GIMS for genotype clusters.

LLR (log-likelihood ratio): A measure of the geographic concentration of a specific genotype in a county, compared with the national distribution of that same genotype, throughout a 3-year period. The higher the LLR, the greater the evidence that the local genotype cluster within the county represents a greater geospatial concentration than the national average, which might indicate recent transmission of *M. tuberculosis*.

MDR: Multidrug-resistant (MDR) tuberculosis strains are resistant to at least isoniazid and rifampin.

MIRU-VNTR: Mycobacterial interspersed repetitive unit–variable number tandem repeat typing analysis. MIRU-VNTR is a polymerase chain reaction (PCR)-based genotyping assay. The CDC genotyping program performs 24-locus MIRU-VNTR analysis on every isolate submitted for genotyping. Before 2009, only 12-locus MIRU-VNTR was performed.

***Mycobacterium bovis*:** A member of the *M. tuberculosis* complex that is commonly associated with cattle, particularly in countries with a low socioeconomic status. In the United States, human cases of *M. bovis* TB typically have a foodborne origin (e.g., consumption of unpasteurized dairy products). *M. bovis* is typically resistant to pyrazinamide. Identification of TB isolates that are *M. bovis* can be performed through genotyping; however, this information should not be relied on for clinical decision making.

***Mycobacterium tuberculosis* complex:** Often abbreviated *MTC*, a group of closely related mycobacterial species that can cause latent TB infection (LTBI) and TB disease (i.e., *M. tuberculosis*, *M. bovis*, *M. bovis* bacillus Calmette-Guérin, *M. africanum*, *M. canetti*, *M. microti*, *M. pinnipeditii*, and *M. mungi*). Among humans, the majority of TB cases are caused by *M. tuberculosis*.

NTGS: The National TB Genotyping Service has provided TB genotyping services to local and state TB control programs since 2004. National genotyping laboratories are contracted by CDC to provide genotyping services at no cost to patients, health care providers, or health departments.

NTSS: National TB Surveillance System administered by CDC. NTSS collects surveillance data through an electronic reporting registry. Data collected include sociodemographic, clinical, and risk factor variables that are reported to CDC by states and local health departments.

PCRType: A designation for each unique combination of spoligotype and 12-locus MIRU–VNTR results. PCRType is designated as *PCR* followed by five digits, which are assigned sequentially to every genotype identified in the United States (e.g., PCR01974).

Polymerase chain reaction (PCR): A laboratory method that can rapidly amplify limited quantities of DNA, thereby enabling certain types of laboratory testing. The national genotyping laboratories routinely use two PCR-based techniques, spoligotyping and MIRU-VNTR analysis.

Recent transmission: Although the precise time interval is not well-defined, recent transmission for TB is often considered to be TB disease that is attributable to exposure 2–3 years before disease onset. That is, the chain of transmission spanning from exposure to source case through onset of symptoms for secondary cases is <3 years. Immunocompromised patients (e.g., patients with human immunodeficiency virus infection or diabetes) might be at a higher risk for acquiring TB disease.

Relapse versus reinfection: A case of relapsed TB represents a worsening of signs and symptoms of disease after a period of improvement, caused by the same strain of *M. tuberculosis*. TB that represents a new infection (or reinfection) is disease caused by a second infection (often with a strain different from the strain that caused the initial infection). Genotyping the initial and the subsequent *M. tuberculosis* isolate might distinguish these two possibilities.

Report of a Verified Case of TB (RVCT): National surveillance data on patients with tuberculosis is recorded on this form and subsequently reported to CDC's National TB Surveillance System.

Restriction fragment length polymorphism (RFLP): Also called IS6110-based, RFLP analysis was the first widely used method for genotyping *M. tuberculosis* isolates. A genotyping technique based on measuring the number and length of specific DNA fragments that are cut by using specific restriction enzymes.

Spoligotyping: Spacer oligonucleotide genotyping. A genotyping technique based on spacer sequences located in the direct repeat region in the chromosomes (genetic makeup) of the *M. tuberculosis* complex. The spoligotype is reported as a 15-digit number.

References

1. Cowan LS, Crawford JT. Genotype analysis of *Mycobacterium tuberculosis* isolates from a sentinel surveillance population. Emerg Infect Dis 2002;8:1294–302.
2. Haddad MB, Diem MA, Cowan LS, et al. Tuberculosis genotyping in six low-incidence states, 2000–2003. Am J Prev Med 2007;32:239–43.
3. Ghosh S, Moonan PK, Cowan L, Grant J, Kammerer S, Navin TR. Tuberculosis Genotyping Information Management System: enhancing tuberculosis surveillance in the United States. Infect Genet Evol 2012;12:782–8.
4. Centers for Disease Control and Prevention (CDC). (2010, March 19). Monitoring Tuberculosis Programs---National Tuberculosis Indicator Project, United States, 2002–2008. MMWR. Morbidity and Mortality Weekly Reports. Retrieved from <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5910a3.htm>