

# 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 a variety of ways. These applications are summarized as follows:

#### **Patient-level Applications of Genotyping**

- Complete contact investigations
  - Confirm or refute patient connections (epidemiologic linkages) identified that may or may not be found 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 cases with recurrent TB disease

#### **Population-level Applications of Genotyping**

- Detect potential outbreaks using geospatial or other analyses of genotype clusters
- Refute outbreaks when cases thought to be part of the same outbreak have non-matching 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 get 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 (NTG-SN), a 5-year initiative which established the utility of genotyping in TB control efforts.<sup>1</sup> In 2004, based on the knowledge gained from NTG-SN and associated studies,<sup>2</sup> CDC established the National TB Genotyping Service (NTGS) and funded a national genotyping laboratories, located in Michigan, to genotype at least one *M. tuberculosis* isolate from each culture-positive TB case reported in the United States.<sup>3</sup> All TB control programs may use NTGS at no cost to the patients, healthcare 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, over 105,000 *M. tuberculosis* isolates have been successfully genotyped through NTGS and its partnerships between CDC, national genotyping laboratories, and 58 states and jurisdictions.

In 2010, CDC launched the TB Genotyping Information Management System (TB GIMS), a secure web-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 on genotypes and clusters, data quality checks, aggregate reports, maps, and outbreak detection tools. TB GIMS currently has over 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 the use of geospatial analysis to identify unusual groupings of TB cases with matching genotypes that may 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.

Currently, CDC's primary outbreak detection method is based on identifying higher than expected geospatial concentrations of a TB genotype in a specific county, compared to 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. LLR is then classified into alert levels within TB GIMS based on established cut points. Clusters are classified as no alert (LLR < 5.0), medium alert (LLR  $\geq$  5.0 and < 10.0), or high alert ( $\geq$  10.0). The alert level and changes in alert levels (e.g., from none to medium or high) can help TB programs identify outbreaks and prioritize TB genotype clusters for further investigation or intervention.

---

<sup>1</sup> Cowan LS, Crawford JT. Genotype analysis of *Mycobacterium tuberculosis* isolates from a sentinel surveillance population. *Emerg Infect Dis* 2002; 8(11): 1294–302.

<sup>2</sup> Haddad MB, Diem MA, Cowan LS, et al. Tuberculosis genotyping in six low-incidence states, 2000–2003. *Am J Prev Med* 2007; 32(3):239–43.

<sup>3</sup> 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.

## **Genotyping Terminology**

In NTGS, a genotype is currently 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 “GENType” designated as “G” followed by five digits, which are assigned sequentially to every genotype identified in the U.S. (e.g., G00162). This nomenclature is designed for convenience and ease of communication, but the specific numbers assigned have no additional significance outside of 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 the entire country. Additionally, since 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 (NTIP) national genotyping surveillance coverage objective is 94%.

## **GLOSSARY**

### **Alert level**

A mechanism used by TB GIMS to notify users of genotype clusters, possibly representing TB out-breaks, in a specific county. The alert level is determined by the log likelihood ratio statistic (LLR) for a given cluster. This is calculated by TB GIMS and is updated whenever a new case is added to a genotype cluster. Email notifications are generated whenever an alert level changes from a “none” LLR (0–5) to “medium” LLR (5.1–10) or “high” LLR (>10), 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 may 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* could have been transmitted between them. Patients that named each other as contacts have an epidemiologic link. However, an epidemiologic link could be a location where the two persons spent time together or an activity that brought them together.

### **Geospatial concentration**

Geospatial concentration is a measure of how concentrated a genotype is in time and space. It suggests that recent transmission has occurred since cases with the same genotype in the same location are more likely to have come in contact with each other. TB GIMS uses the log likelihood ratio (LLR) to generate a numeric measure of geospatial concentration of a given TB genotype.

### **Genotype**

The designation that represents one or more of the three genotyping techniques used for *M. tuberculosis*: spoligotyping, MIRU-VNTR analysis, and IS6110-based RFLP. These designations were developed to facilitate communication of genotyping information within and between TB programs. In the U.S., we use GENType or PCRTType to define a genotype.

**Genotyping cluster**

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

**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 U.S. (e.g., G00017).

**LLR (log likelihood ratio)**

A measure of the geographic concentration of a specific genotype in a county, compared to the national distribution of that same genotype, over 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*.

**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 to ensure that demographic, risk factor and geographic data can be viewed in TB GIMS for genotype clusters.

**MDR**

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

**MIRU-VNTR**

Mycobacterial interspersed repetitive unit–variable number tandem repeat typing analysis. MIRU-VNTR is a PCR-based genotyping assay. The CDC genotyping program currently 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 the developing world. In the United States, human cases of *M. bovis* TB generally have a foodborne origin, such as through consumption of unpasteurized dairy products. *M. bovis* is typically resistant to pyrazinamide (PZA). Identification of TB isolates that are *M. bovis* can be done 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* BCG, *M. africanum*, *M. canetti*, *M. microti*, *M. pinnipedii*, and *M. mungi*). In humans, most TB is caused by *M. tuberculosis*.

**NTGS**

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

**NTSS**

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

**PCR**

Polymerase chain reaction (PCR) is a laboratory method that can rapidly amplify small 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.

**PCRType**

A designation for each a 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 U.S. (e.g., PCR01974).

**Recent Transmission**

Although the precise time interval is not well defined, “recent” transmission for TB is often considered to be TB disease that is due to exposure 2-3 years prior to disease onset. That is, the chain of transmission spanning from exposure to source case through onset of symptoms for secondary cases would be <3 years. Immunocompromised patients (e.g., patients with HIV or diabetes) may be at a higher risk for acquiring TB disease.

**Relapse vs. 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 that is different from the strain that caused the initial infection). Genotyping the initial and the subsequent *M. tuberculosis* isolate might distinguish these two possibilities.

**RFLP**

Restriction fragment length polymorphism. Also called IS6110-based restriction fragment length polymorphism (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 using specific restriction enzymes.

**RVCT**

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

**Spoligotyping**

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