USING WHOLE-GENOME SEQUENCING AND SOCIAL MEDIA TO ENHANCE A TUBERCULOSIS OUTBREAK INVESTIGATION

Jillian Knorr, MPH; Herns Modestil, BS; Shama D. Ahuja, PhD, MPH; Jeanne Sullivan Meissner, MPH

Bureau of Tuberculosis Control, New York City Department of Health and Mental Hygiene, Queens, New York

NYC Health

BACKGROUND

- In 2001, the New York City (NYC) Bureau of Tuberculosis Control (BTBC) implemented universal genotyping for all culture-positive tuberculosis (TB) cases using spacer oligonucleotide typing (spoligotyping) and IS6110 restriction fragment length polymorphism (RFLP) analysis
  - RFLP was discontinued in 2017
- The Centers for Disease Control and Prevention (CDC) began universal genotyping for the United States in 2004 using spoligotyping and 12-loci mycobacterial interspersed repetitive unit–variable number tandem repeat (MIRU) analysis
  - 24-loci MIRU was introduced in 2009

CLUSTER INVESTIGATION IN NYC

- Genotype clusters are defined as two or more cases with matching genotyping results across at least two genotyping methods; definitions have changed over time based on availability of results from different genotyping methods
- Clustered cases are reviewed, prioritized and assigned for epidemiologic investigation based on factors including cluster growth, patient characteristics and evidence of recent, local TB transmission

WHOLE-GENOME SEQUENCING

- Whole-genome sequencing (WGS) for large TB outbreaks has been available through the CDC since 2013; out-of-jurisdiction results are available by request given jurisdiction permission
- The availability of WGS results, including high-quality single nucleotide polymorphism (SNP) analysis and phylogenetic trees, has enabled further differentiation of TB strains and informs cluster detection, prioritization and investigation
  - Universal SNP thresholds for interpreting WGS results have not been formally established; however, 0-5 SNP differences between isolates are generally considered to be suggestive of recent transmission

OUTBREAK BACKGROUND

- In 2012, BTBC identified a quickly-growing TB outbreak within an endemic strain (Cluster A) first identified in NYC in 1997; few known links existed among outbreak patients

METHODS

- BTBC requested high-quality SNP analysis for all NYC cases with similar or exact-match spoligotype and MIRU-24 results identified since 2012

OUTBREAK INVESTIGATION OBJECTIVES

1. Identify potentially-exposed individuals and sites of transmission and ensure appropriate evaluation and treatment for all exposed individuals
2. Identify possible epidemiologic links among Cluster A cases
3. Investigate possible links to interjurisdictional cases and potential for cross-jurisdictional transmission
4. Conduct active surveillance of genotyping results to rapidly detect new outbreak cases

OUTBREAK INVESTIGATION ACTIVITIES

- Analysis of existing demographic, clinical, social, geographic, and genotype data
- Review of contact investigation and case management notes
- Medical chart review
- Patient re-interview using an open-ended structured questionnaire
- Queries of non-BTBC medical record, public record, and social service databases

INTERPRETING WGS RESULTS

- Results of high-quality SNP analysis were examined to determine which patients may be linked by recent transmission based on SNP distances
- Isolates with similar results (i.e., exact matches or ≤5 SNPs) were considered outbreak cases

USE OF SOCIAL MEDIA

- BTBC staff conducted searches of social media platforms for outbreak patients and their contacts
  - When available, social media profiles and publically-available data were reviewed to establish potential links between patients
    - Available information included aliases/nicknames, photos, posts, comments, and "friends"
  - Patients were never asked directly about anything found on their social media accounts; information was used to establish links between patients and guide interview questions

RESULTS

Figure 1: Number of Cluster A cases[1] by year verified as a tuberculosis case and out-of-jurisdiction status, 1997-2018 (n=103)
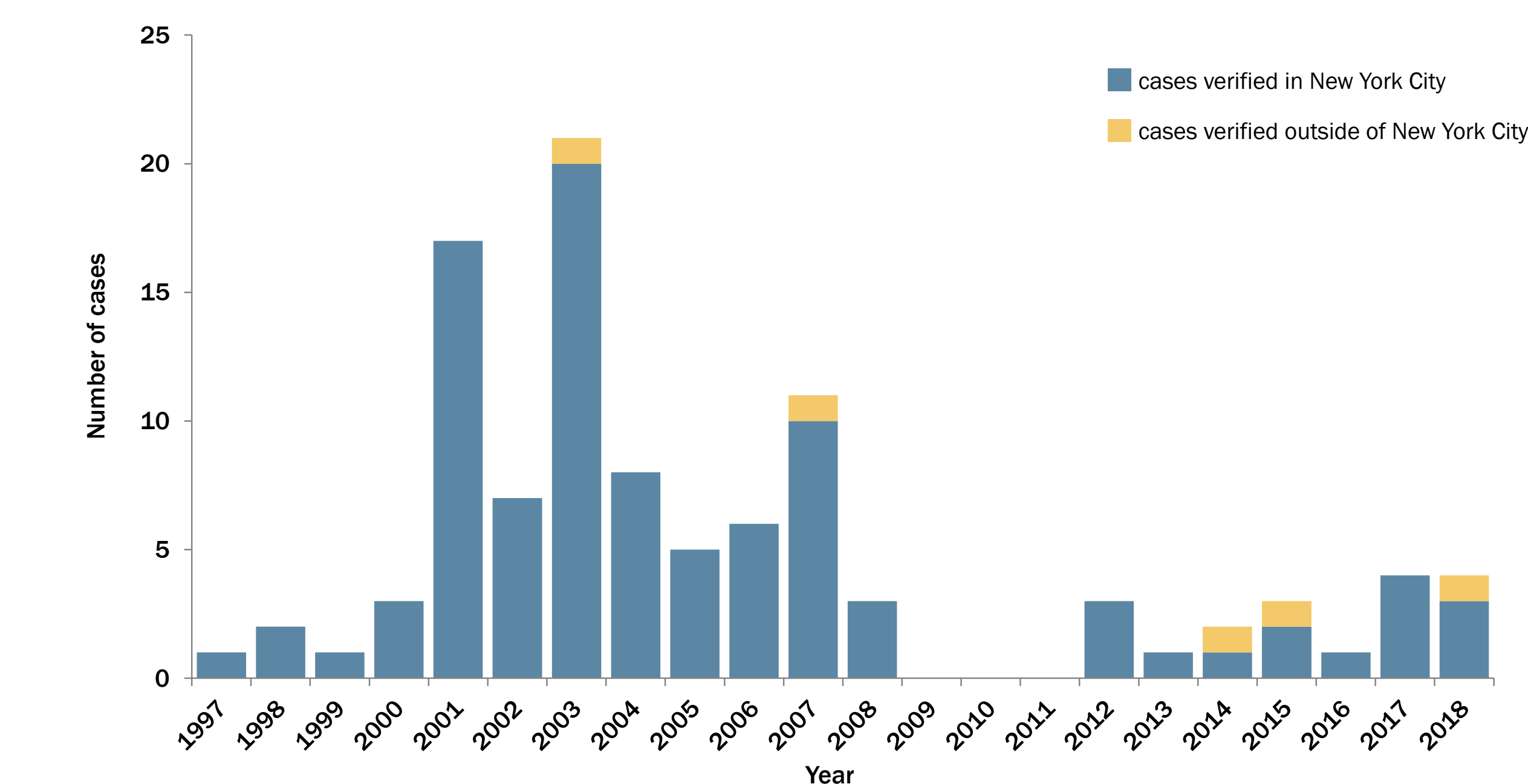
1. Cases were defined as clustered based on matching spoligotype and RFLP pattern, or exact-match or similar spoligotype and MIRU-24 pattern

Table 1: Select demographic, clinical, and social characteristics of Cluster A patients by year verified as a tuberculosis case and outbreak status

Patient Characteristics	Cases verified from 1997-2008 (n=85)	Non–outbreak cases verified from 2012-2018 (n=8)	Outbreak cases verified from 2012-2018 (n=10)
Male	58 (68%)	4 (50%)	8 (80%)
Median age (range)	43 (18-77)	55 (16-76)	29 (19-59)
Birth in the United States	61 (72%)	5 (63%)	10 (100%)
Non-Hispanic Black race/ethnicity[1]	50 (82%)	4 (80%)	7 (70%)
Pulmonary site of disease	78 (92%)	6 (75%)	9 (90%)
Human immunodeficiency virus (HIV) infection[2]	56 (68%)	3 (38%)	0 (0%)
Residence in Manhattan	43 (51%)	4 (50%)	7 (70%)
History of substance abuse	57 (67%)	3 (38%)	8 (80%)

1. Race/Ethnicity is among patients born in the United States
2. Human immunodeficiency virus (HIV) infection is among those with known HIV status

- A large proportion (68%) of cluster cases verified from 1997-2008 were among patients with human immunodeficiency virus (HIV) infection; Cluster investigation at that time identified a single-room occupancy hotel as a potential site of TB transmission
- In 2014, there was a notable shift in patient characteristics, with more cases among young men with history of marijuana use and no known HIV infection
- 80% of outbreak patients were linked, either by residence at diagnosis or residence of a known associate, to the same multi-unit apartment complex (Complex Y) in Manhattan

Figure 2: Whole-genome sequencing (WGS) results for cluster A cases, 2012-2018[1,2,3,4] (n=16)
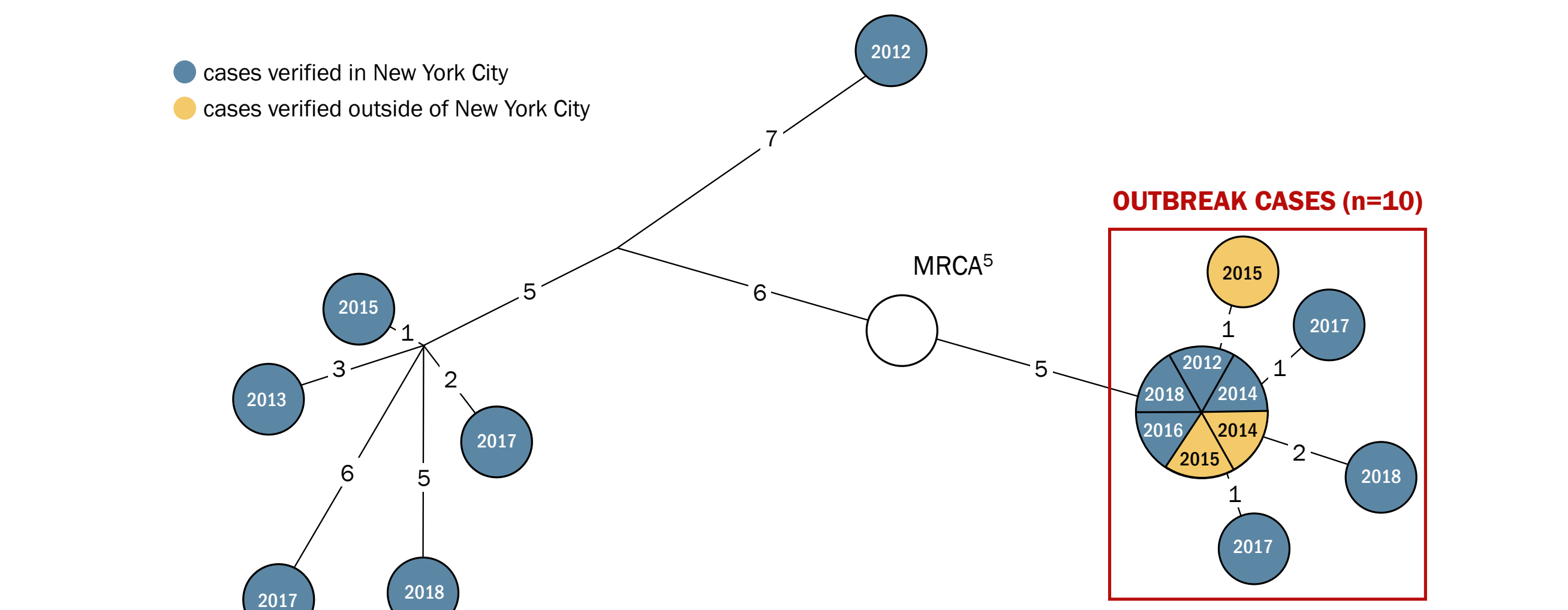
1. Whole-genome sequencing (WGS) results were unavailable for one 2012 case and pending for one 2018 case
2. Circular nodes represent individual patient isolates; isolates with identical WGS results are shown together in one node
3. Nodes are labeled with the year in which each patient became a verified tuberculosis case
4. Lines connecting each node are labeled with the single nucleotide polymorphism (SNP) distance between isolates
5. The central white 'MRCA' node is the 'most recent common ancestor' and does not represent any patient's isolate but rather a hypothetical genome from which all isolates on the tree are descended

- WGS identified 10 outbreak cases likely involved in recent transmission
- Three outbreak cases were verified by jurisdictions outside of NYC

Figure 3: Diagram of known epidemiologic links among Cluster A outbreak patients[1], 2012-2018 (n=8)
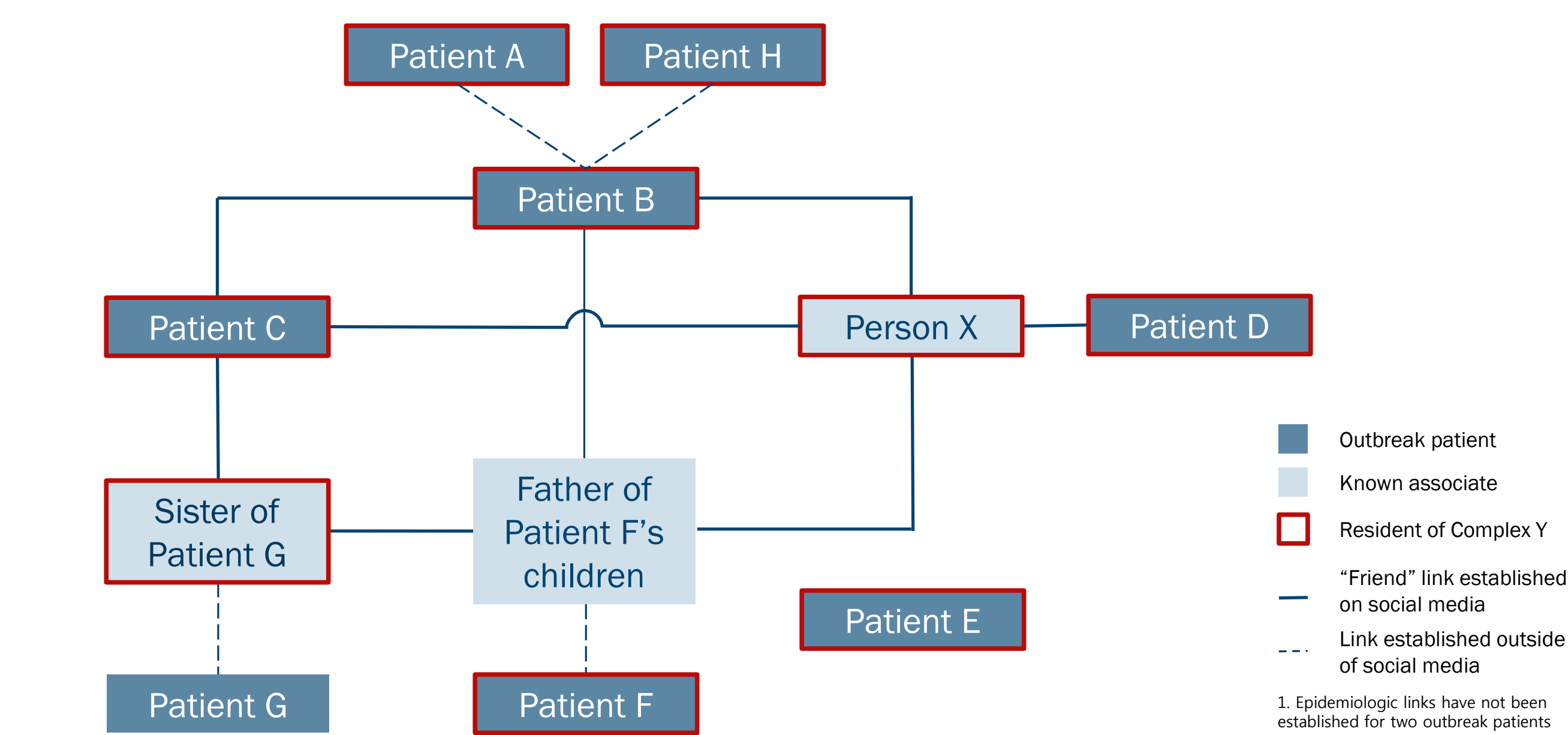
EPIDEMIOLOGIC LINKS

- Patients A, B, and H had known epidemiologic links prior to investigation; all were members of the same extended family and resided in Complex Y
- Social media searches revealed Patient B was a "friend" of Patient C. Patients B, C, and D shared a common "friend," Person X. Patient's C and D also resided in Complex Y
- Patient G's sister resided in Complex Y, and Patient G reported spending time with her there; Patient G's sister was a "friend" of Patient C
- Patient F was a resident of Complex Y; social service database searches identified the father of Patient F's children, who was a "friend" of Patient B, Patient G's sister, and Person X
- Patient E was a resident of Complex Y but was not linked through social media to any other outbreak patient

DISCUSSION

- WGS results helped to focus outbreak investigation efforts by identifying 10 (56%) cases likely linked by transmission among 18 Cluster A cases verified since 2012; results also supported outbreak investigation findings and known epidemiologic links
- WGS results refuted transmission between outbreak cases and other recently verified Cluster A cases with no established epidemiologic links, including two 2017 cases and one 2018 case
- WGS enabled the identification of potential cross-jurisdictional TB transmission
- While genotyping aids in identifying and assessing transmission, epidemiologic investigation and case management are needed to give results context and inform interventions
- Epidemiologic investigation combined with WGS and social media helped uncover a social network among outbreak patients
- Identifying social contacts through traditional contact investigation and patient interview was unsuccessful in this group of mostly young men
- Social media was an invaluable tool for elucidating relationships and identifying social network links between patients
- Additional research is still warranted to better understand TB mutation rates and formalize WGS SNP thresholds for the detection of recent transmission and to establish best practices for clustering using WGS in the absence of other conventional genotyping methods