

## The National Tuberculosis Archive: Briefing Document

### Proposal for the nation's first integrated, comprehensive, biological and information resource for a human pathogen

Tuberculosis (TB) infects approximately one third of the world's population, with 8.8 million new cases and 2-3 million deaths per year. In the United States, there are approximately 13,000 new cases each year. TB is expensive. A recent study published in the *New England Journal of Medicine* estimates that the United States will spend approximately \$2 billion over the next 20 years in treating just immigrants from Mexico<sup>1</sup>. But disturbingly, new strains of TB are emerging that are resistant to conventional antimicrobial therapies. Treating MDR-TB (multi-drug resistant) TB costs approximately \$200,000 per patient in industrialized nations, while XDR-TB (extensively drug resistant) is essentially untreatable given standard drug regimes<sup>2</sup>.

Today, we have no resource that systematically tracks the genetic, clinical, epidemiological, and phylogenetic data for any pathogen. For TB, even though much of this data both exists and is collected, it is dispersed and unavailable for on-demand interrogation for evidenced-based decision-making. To address this clear gap and vulnerability, a team of top TB and informatic experts from the nation's top universities, private research and hospital organizations, as well as CDC TB experts, developed and published a plan: writing in the journal *Science*<sup>3</sup>, the team proposed the establishment of the nation's first integrated, comprehensive, biological and information resource for tuberculosis.

The proposed National Tuberculosis Archive would be the first integrated, comprehensive, biological and information resource for a major infectious disease. In cooperation with CDC, the Archive would contain a sample from every verified case of TB in the United States along with each isolate's integrated clinical, epidemiological, genomic, and phylogenetic data. The resource would be web-accessible for secured personnel, allowing cross-data type interrogation linked with the underlying bar-coded isolates for targeted investigations. This integrated resource would, by its existence, create the world's premier integrated data set for any infectious disease, and be an important pilot model for other infectious diseases. The Archive would be a milestone in 21<sup>st</sup> century, large scale scientific approaches to world-wide infectious diseases.

Implementation is estimated to cost \$15 million over three years. The relative low cost for the high return on investment is possible because key data is already being collected under the nation's public health system. CDC is already handling isolates as part of its Universal Genotyping Program. What is needed is an informatics and archiving approach to organize and present these assets in a manner responsive to national disease preparedness. Accelerating the National Archive's deployment, the Archive would not in-and-of itself link to patient data, though that link would be available to authorities. This would establish the Archive as a pathogen-centric resource, compliant with patient privacy concerns, yet central to infectious disease monitoring and prediction.

For more information please contact Dr. Damian Gessler, National Center for Genome Resources, 2985 Rodeo Park Drive East, Santa Fe, NM 87505 (505) 995-4403 ddg@ncgr.org

---

<sup>1</sup> K. Schwartzman, O. Oxlade, R. G. Barr, F. Grimand, I. Acosta, *et al.* *N. Engl. J. Med.* **353**,1008 (2005).

<sup>2</sup> In Speaker's case, the strain is resistant to at least 10 of 14 drugs. Still, doctors are optimistic about his long-term prognosis, in part because he is a good candidate for surgery.

<sup>3</sup> Gessler D, Dye C, Farmer P, Murray M, Navin T, Reves R, Shinnick T, Small PM, Yates T, Simpson G. (2006) Public health. A National Tuberculosis Archive. *Science* **311**(5765): 1245-6. PMID: 16513968

## Frequently Asked Questions

- **What is it?** The National Tuberculosis Archive is proposed as the nation's first integrated, comprehensive, biological and informatic information resource for a human pathogen. It is a system that archives culture-positive tuberculosis (TB) isolates (bacterial samples from a patient) from every verified case of TB in the United States, and associates these isolates with rigorous clinical and epidemiological data as collected by the official RVCT (Report of Verified Case of Tuberculosis). Additionally, the CDC has already embarked on overseeing the genotyping of every isolate. Genotyping is a process similar to sequencing, yet instead of determining the actual genetic code of the organism, it determines a summary "genetic fingerprint" that allows isolates to be identified and grouped by genetic relatedness. Genotyping is of lower resolution—less information—than sequencing. This information can be used to build a phylogenetic tree—a tree of genetic relatedness among isolates.

The Archive is proposed as an integrated biological and information resource: as such it bar-codes and archives the biological isolates, and saves the clinical, epidemiological, and genomic data in a credentialed access, web-accessible database.

- Researchers will be able to search on metadata, such as multi-drug resistance, and discover how those isolates are related to other isolates with other traits, such as high transmission rates. They can then request those isolates from the Archive and target their experiments more directly on isolates of clinical or epidemiological concern;
  - Clinicians will be able to query the database to see how other patients with similar infections responded to drug treatments. Individualized, patient-identifiable data will not be available, but clinicians will be able to use their normal clinical and public health channels to further their inquiries.
  - Public health officials will be able to track genetic signatures across jurisdictions and over time—something that is still surprisingly difficult to do today. This is aimed directly at building a better rapid-response, evidence-based infrastructure for decision-makers.
- **What's the larger picture?** Science, medicine, and public health form the triad of our health delivery system. In an ideal world, scientific discoveries are translated into medical advances, both of which are brought to bear for the betterment of the common good via public health. But in the real world substantial technological, sociological, and fiscal considerations impede the timely translation of research into benefits for healthier citizens, clinical practice, and policy formation. The National Tuberculosis Archive integrates across science, medicine, and public health to better accelerate this health delivery triad.
  - **Why tuberculosis?** Three reasons: 1) Relevancy; approximately two billion people in the world have TB; there are 8.8 million new cases each year, resulting in 2-3 million deaths. TB has become known as the "Face of AIDS" because of its tendency to activate in immunocompromised patients, thereby making it the most frequent HIV-associated opportunistic disease. Indeed, worldwide, of the people diagnosed with HIV/AIDS, more will eventually die from tuberculosis than will die from any single other complication associated with AIDS. This is despite the fact that AIDS is expensive and difficult to treat, while non-drug resistant TB is readily curable with antimicrobials; 2) Superb infrastructure; no other disease has the type of extensive and detailed clinical and epidemiological data collection infrastructure already established for tuberculosis. Clinical, epidemiological, and genotypic data is being collected on every reported case, meaning that we can get a full population assessment of this disease. This is unusual, because for most diseases we usually see a much smaller sampling of the pathogen from the population at large. Genetically, the disease-causing bacterium *Mycobacterium tuberculosis* has been sequenced, so important base-line genetic data is available; 3) Manageability; there are approximately 13,000 cases/yr in the US, though handling these cases is expected to cost well over than \$2 billion dollars over the next 20 years. Techniques to collect and archive 13,000 isolates/yr are well known.

The combined factors of relevancy, data infrastructure, and manageability mean that the National Tuberculosis Archive is posed to provide the world's premier data set of any human pathogen. Alternatively, not integrating this currently disparate data by not creating the National Tuberculosis Archive would leave vast amounts of data under-utilized.

- **What is the value-add to what is already out there?** CDC has already implemented its Universal Genotyping program to genotype one or more isolates from every culture-positive case of tuberculosis in the United States. Yet this data needs to be integrated with clinical, epidemiological, and phylogenetic data to extract its full value. The National Institutes of Health is funding researchers to study TB. But no one is yet putting this together to make a national biological and informatic resource center directly aimed at building an evidence-based infrastructure for researchers, clinicians, and public health policy makers.
- **Is archiving safe?** Yes. With two billion people in the world currently harboring tuberculosis, access to the bacterium is as easy as offering a handkerchief. So archiving isolates that are characterized for specific clinical, epidemiological, and genomic data actually increases security by auditing and controlling access to those specific strains relevant to research and public health. The National Tuberculosis Archive is proposed to secure information in two complementary ways. First, it will require credentialed access according to how data is currently handled in the public domain: data that is publicly available will remain publicly available, while other data will require registration and credentialed access. Second, underlying patient-identifiable data will not be stored in the system. The National Tuberculosis Archive will coordinate with state public health departments and CDC to store summarized and aggregated data. No one will be able to access patient-identifiable data because it will never be stored in the system.
- **Who does it help?** The National Tuberculosis Archive gives researchers, clinicians, and public health policy makers an evidence-based infrastructure to better drive their decision making. This helps people with tuberculosis, as it also helps the general public from endangerment to a rise in TB as experienced in the 1990s, and increases in MDR-TB and XDR-TB experienced today. Tuberculosis by itself is not a biothreat. But establishing an integrated biological and information resource for a major world-wide infectious disease will give us valuable knowledge in how to expand the model for more acutely dangerous pathogens, such as influenza and biothreat agents.
- **How much will it cost?** We estimate \$15 million over three years.
- **When could it be done?** All the required know-how and enabling technologies exist today. The Archive could be operational within three years of the start of funding.
- **What are the publication details?** Gessler D, Dye C, Farmer P, Murray M, Navin T, Reves R, Shinnick T, Small PM, Yates T, Simpson G. (2006) Public health. A National Tuberculosis Archive. *Science* **311**(5765): 1245-6. PMID: 16513968

Authors:

- Damian Gessler, Ph.D. Program Lead, National Center for Genome Resources, Santa Fe, NM 87505, USA ([www.ncgr.org/staff/bios#ddg](http://www.ncgr.org/staff/bios#ddg)).
- Christopher Dye, Ph.D. Coordinator, Tuberculosis Monitoring and Evaluation, Stop TB, World Health Organization, Geneva 27, Switzerland ([www.who.int/tb/about/dye\\_biodata/en/index.html](http://www.who.int/tb/about/dye_biodata/en/index.html)).
- Paul Farmer, M.D., Ph.D.
  - Department of Social Medicine, Harvard Medical School, Boston, MA 02115 ([www.hms.harvard.edu/dsm/WorkFiles/html/people/faculty/PaulFarmer.html](http://www.hms.harvard.edu/dsm/WorkFiles/html/people/faculty/PaulFarmer.html)).
  - Founding Director of Partners in Health ([www.pih.org/index.html](http://www.pih.org/index.html)).
- Megan Murray, M.D., MPH, DPH

- Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115 ([www.hsph.harvard.edu/faculty/MeganMurray.html](http://www.hsph.harvard.edu/faculty/MeganMurray.html)).
  - Department of Medicine, Harvard Medical School, Boston, MA 02115, USA (<http://hms.harvard.edu/WhitePagesPublic.asp?task=showperson&id=179271374171271371179276&a=hms&r=12&kw=>).
  - Infectious Disease Unit, Massachusetts General Hospital ([www.mgh.harvard.edu/id/faculty/research\\_interests/?ID=23](http://www.mgh.harvard.edu/id/faculty/research_interests/?ID=23)).
  - Social Medicine Health and Inequalities, Brigham and Women's Hospital ([www.brighamandwomens.org/socialmedicine/murraybio.asp](http://www.brighamandwomens.org/socialmedicine/murraybio.asp)).
- Thomas Navin, M.D., Chief, Surveillance & Epidemiology Branch, Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta GA 30333 ([www.cdc.gov/nchstp/tb/default.htm](http://www.cdc.gov/nchstp/tb/default.htm)); U.S. Department of Health and Human Services, Washington, D.C. 20201 ([www.hhs.gov](http://www.hhs.gov)).
- Reves, Randall, M.D., M.Sc.,
    - Denver Public Health, Denver Health and Hospital Authority, Denver, CO 80204 ([www.denverhealth.org/Hospital/PhysicianDirectoryDetail.aspx?id=256](http://www.denverhealth.org/Hospital/PhysicianDirectoryDetail.aspx?id=256)).
    - Division of Infectious Diseases, University of Colorado Health Sciences Center, 4250 East 8<sup>th</sup> Avenue, Room 1813, Campus Box B168, Denver, CO 80220 ([www.uchsc.edu/id/faculty/faculty.htm](http://www.uchsc.edu/id/faculty/faculty.htm)).
- Shinnick, Thomas, Ph.D., Chief, TB/Microbacteriology Branch, Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta GA 30333 ([www.cdc.gov/nchstp/tb/default.htm](http://www.cdc.gov/nchstp/tb/default.htm)); U.S. Department of Health and Human Services, Washington, D.C. 20201 ([www.hhs.gov](http://www.hhs.gov)).
- Small, Peter M.D.
    - Institute for Systems Biology, Seattle, WA, 98103 ([www.systemsbiology.org/Scientists\\_and\\_Research/Faculty\\_Groups/Small\\_Group](http://www.systemsbiology.org/Scientists_and_Research/Faculty_Groups/Small_Group)).
    - Senior Program Officer for Tuberculosis, Global Health Program, Bill & Melinda Gates Foundation, Seattle, WA 98102 ([www.gatesfoundation.org/GlobalHealth/Pri\\_Diseases/Tuberculosis/default.htm](http://www.gatesfoundation.org/GlobalHealth/Pri_Diseases/Tuberculosis/default.htm)).
- Terry Yates, Ph.D. (Deceased)
    - Vice Provost for Research, University of New Mexico, Albuquerque, NM 87131 (<http://research.unm.edu>).
    - Department of Biology, University of New Mexico, Albuquerque, NM 87131 (<http://biology.unm.edu/yate.htm>).
    - Museum of Southwestern Biology, MSC03 2020, University of New Mexico, Albuquerque, NM, USA 87131-0001 ([www.msb.unm.edu](http://www.msb.unm.edu)).
- Gary Simpson, M.D., Ph.D., MPH, Former Medical Director of Infectious Diseases, New Mexico Department of Health, Santa Fe, NM 87502, USA.