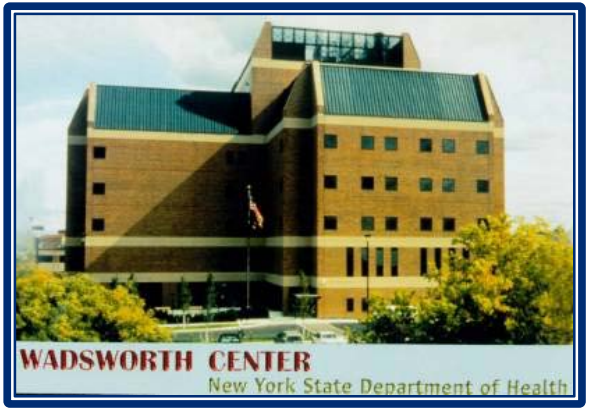
NEW YORK STATE OF OPPORTUNITY. Department of Health | Wadsworth Center

Clinical Testing of *Mycobacterium tuberculosis* by NGS: Two Years Strong



Kimberlee Musser, PhD
Chief, Bacterial Diseases
Wadsworth Center

Why NGS on TB?

Roughly one third of the world's population is infected with TB

2016: 10.4 million new infections, **1.7** million deaths

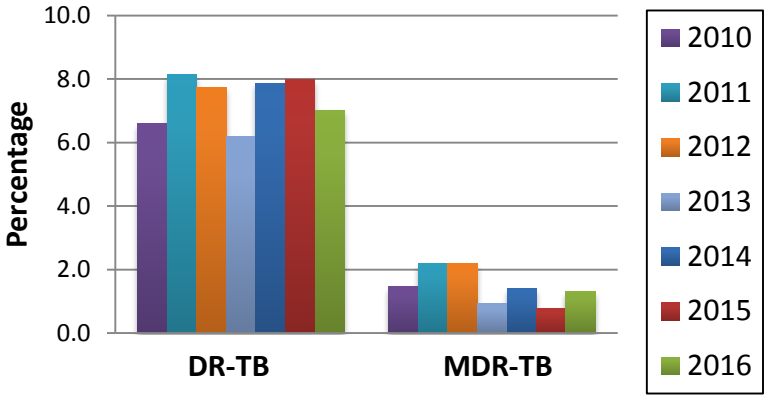
Tremendous infectious control and patient treatment benefit to knowing susceptibili[illegible] to
[illegible]gs and
[illegible]tednes
[illegible]er case

Culture and susceptibility testing can take weeks to months…

Many culture and molecular tests are typically performed utilizing different reagents, instruments and

~800 cases/ year in NY Many drug resistant cases

TB in New York



Universal FAST TRACK program:

- **Implemented in 1993**
- **Rapid detection of MTBC from a priority group of highly infectious NY patients with newly diagnosed AFB smear positive sputum**

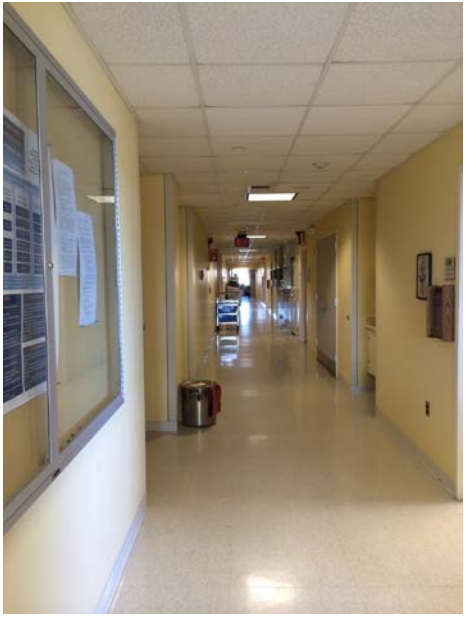
	2010	2011	2012	2013	2014	2015	2016
TB Cases*	954	910	866	872	787	765	768
DR-TB	63	74	67	54	62	61	54
MDR-TB	14	20	19	8	11	6	10
XDR-TB	0	2	2	0	2	0	1

* National rank #3 or #4 each year by number of cases

Whole-genome sequencing for TB?

2013- Wadsworth Center Public Health Genomics Center (PHGC) funding announcement

2014- PHGC funding to test 60 TB isolates by WGS



- **Goals for TB WGS:**
  - **Utilize as soon as possible in testing algorithm to impact patient treatment**
  - **Expand molecular resistance prediction**
  - **Provide more comprehensive results**
    - **mixed infections, heteroresistance, typing**
  - **Assess costs and staff time**

Developing a TB WGS Assay

- Starting material → Day 0 MGIT
- Compare DNA preparation methods
- Nextera XT/ MiSeq
- Build Pipeline
- LIMS/ Epidemiology Reporting (ECLRS)
- Validation Plan

TB Bioinformatics Pipeline

Bioinformatics

Map to Reference Genome

Kraken K-mer matching

Detect spacers

SNP calling with indels

SNP calling ignore indels

MTBC member ID

Spoligotyping

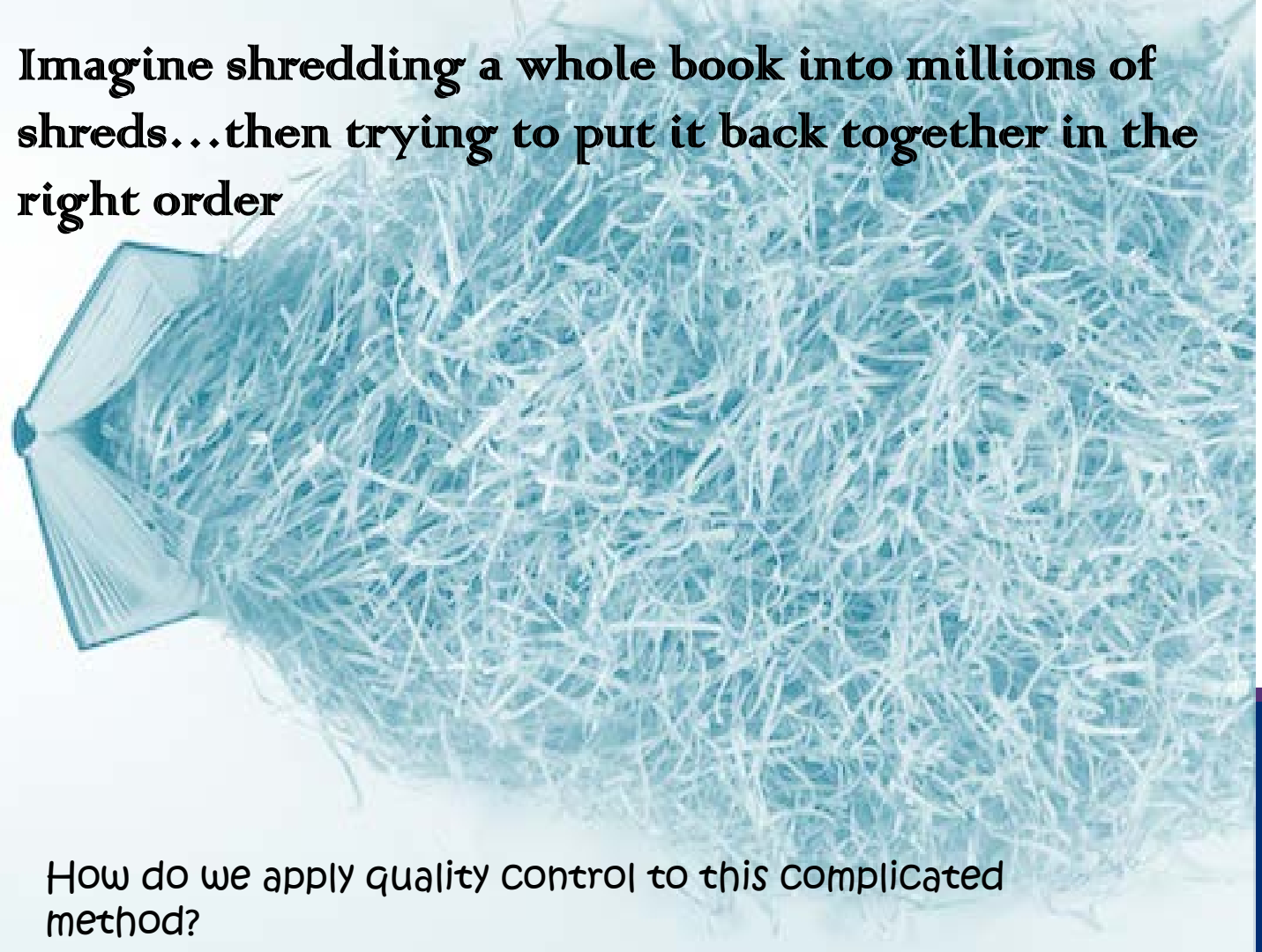
Prediction of antibiotic resistance

Strain typing-relatedness

Report

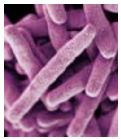
Importing into LIMS

Imagine shredding a whole book into millions of shreds…then trying to put it back together in the right order



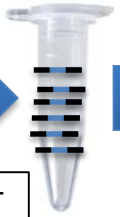
How do we apply quality control to this complicated method?

Controls, QC, and more QC…



Extract DNA

Library Preparation



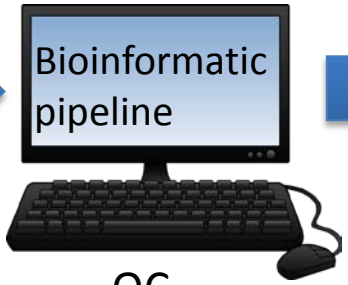
Nextera XT

DNA sequencing

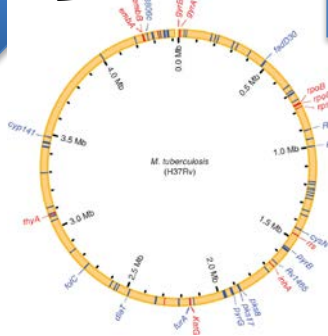
QC

QC

QC

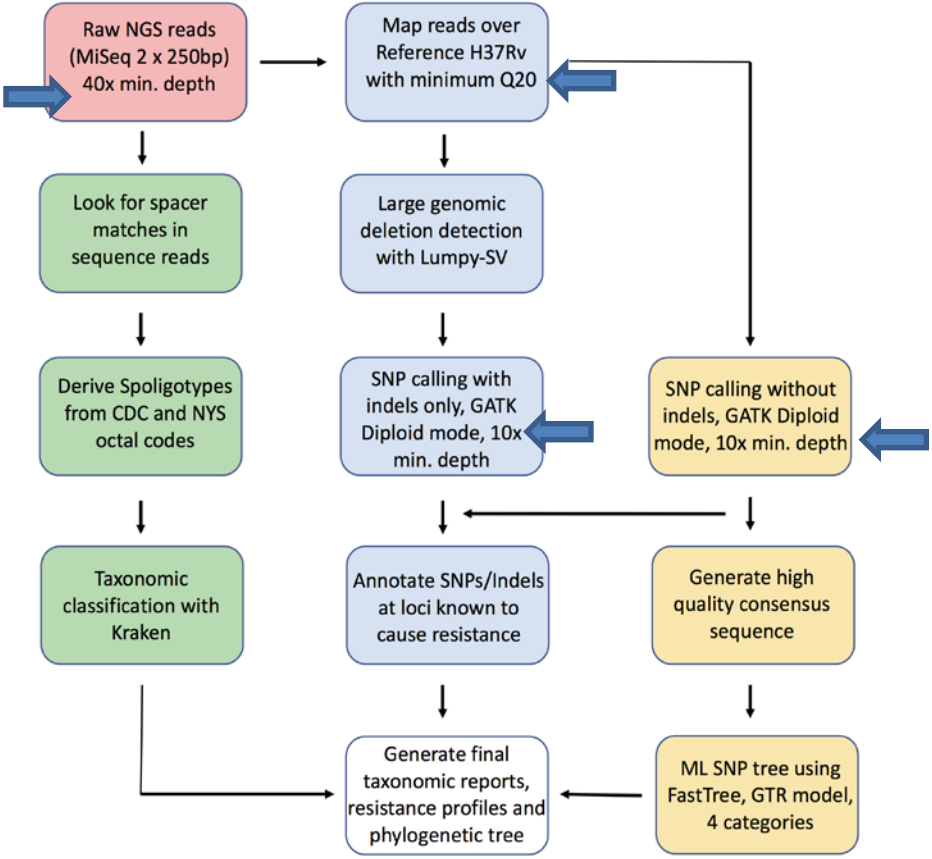


QC



QC

What's in a pipeline?

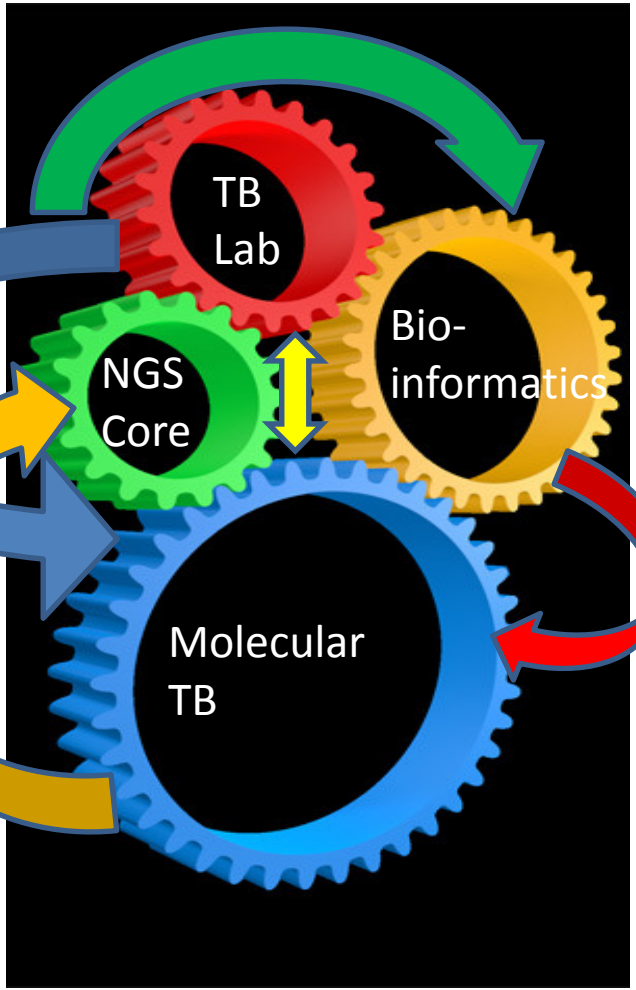


- **Species ID**
- **Spoligotype derivation**
- **Resistance Prediction**
  - **Single Nucleotide Polymorphism (SNP)**
  - **Insertion/ deletions**
- **Phylogenetic Analysis**
- **Results & Reporting**

CLEP QC Guidance

- Minimum base calling of Q20 (99% base call accuracy)
- Minimum average 40x depth of coverage
- All QC metrics must be documented and monitored over time
- All software updates that affect key processes should be revalidated

So many moving parts…



High-confidence mutations

Drug	Locus	Codon/NT position
Rifampin (RIF)	*rpoB*	251, 511, 513, 516, 522, 526, 531, 533, 572
Isoniazid (INH)	*katG* *oxyR-ahpC* promoter region *mabA* promoter region *mabA* *inhA*	279, 315, 394, 525 -81 -17, -15, -8 203 94
Pyrazinamide (PZA)	*pncA/pncA* promoter region	Any nonsynonymous change
Ethambutol (EMB)	*embB*	306, 406, 497
Streptomycin (SM)	*rrs* *rpsL*	512, 513, 516, 906 43, 88
Kanamycin/Amikacin (KAN/AMI)	*rrs*	1401
Kanamycin (KAN)	*eis* promoter region	-10, -37
Fluroquinolones (FLQ)	*gyrA* *gyrB*	74, 90, 91, 94 510
Ethionamide (ETH)	*mabA* promoter region *mabA* *ethA*	-17, -15, -8 203 Frameshift/STOP

Red =Current validated pipeline

Putting it all together…

- Detailed SOP or SOPs
  - All QC, limitations, step by step details
- Reporting
  - Interpretation, disclaimers, examples
- Quality Control
  - Metrics, criteria, controls
- Validation
  - Specificity
  - Reproducibility (Inter- and Intra-)
  - Accuracy verification

The Validation Package…

Please find attached the documentation materials that I have reviewed and approved including:

1. SOPM (SOP MB 91.0.0) for this whole-genome sequencing of Mycobacterium tubercolsis isolates using next generation sequencing technology
2. Validation Package and Appendixes
3. Referenced SOPs from Applied Genomics Technology Core, Mycobacteriology Laboratory, Molecular Bacteriology Laboratory:
   - SOP MB 35.0.0 Employee Training
   - SOP MB 39.0.0 Employee CE
   - SOP MB 53.1.0 MTB complex differentiation
   - SOP AGTC 006.0 Illumina
   - SOP AGTC 007.0 Bioanalyzer
   - SOP AGTC 008.0 Qubit
   - SOP AFB-0016 Processing of Isolates
   - M. tuberculosis heat inactivation protocol
4. Testing reports
5. References for SOP MB 91.1.0 and Validation

Separating SOPs makes future assay development more streamlined

447 Page document

SOP QC Examples…

Post run metrics for assessing success

Unless a run fails entirely, determination of run success will be determined after analysis of the results through the bioinformatics pipeline. The AGTC will monitor standard run metrics to monitor overall performance of the instrument and to look for trends that indicate decreasing performance. These metrics will be will entered into the CMS-MiSeqRunLog.

19.1 Final DAL conc._____pM (Library concentration loaded on instrument)
19.2 Cluster PF (%)_____(Typically >75%)
19.3 Cluster Density_____ Normal (600-1300 K/mm2)
19.4 Q30 %_____ (Typically > 75% - total – all reads combined)
19.5 Reads PF_____ (15 million typical)
19.6 Aligned % (PhiX)_____(~1% expected)
19.7 Error Rate (PhiX)_____(0.6 – 1.8% overall - typical)

C. Compare the results imported into CLIMS with both the CLIMS-report.out (Appendix A) and the identification and resistance report (Appendix B) generated from the pipeline to ensure accuracy.

1. Organism Identification
   a. If Kraken successfully identified the species, the code and species name will be imported and reported out on the CLIMS report.
   b. If no species is reported, the identification can be determined using real-time PCR (See SOP MB 53 "Differentiation of the *Mycobacterium tuberculosis* complex by Real-time PCR"). This result will imported and reported out under the real-time PCR section on the CLIMS report. If the result is inconclusive using real-time PCR, specific genetic features can aid in species identification (REF 1, 2, 3). Go to step c below.
   c. Review the section in the report that is labeled as "All mutations in screened loci" for the following mutations in Table 1 below that may be listed and consult a supervisor to determine the identification. Once the identification is established, use the drop-down menu in CLIMS to fill in the correct code and species name that will appear

Retrospective study: Fluoroquinolone comparison

6. Fluoroquinolone Resistance:

Table 21. Summary of *gyrA/gyrB mutations identified used to predict fluoroquinolone resistance**

High confidence mutations in *gyrA*	Number of isolates found to harbor mutation by WGS during retrospective study	Number of isolates with mutation confirmed by pyrosequencing or Sanger sequencing
Ala90Val	4	4
Ser91Pro	1	1
Asp94Asn	1	1
Asp94Gly	7	7
Total	13	13 (100%)

*No high confidence *gyrB* mutations were identified in this retrospective study

Validating against other molecular tests

Table 22. DST Phenotype results compared to WGS Genotype results for fluoroquinolone resistance (target=*gyrA*)

		Fluoroquinolone DST Phenotype	
WGS Genotype		Resistant	Susceptible
	Resistant	13*	0
	Susceptible	0	60

Validating against DST

Resistance Predictive Value= 100%
Susceptible Predictive value= 100%

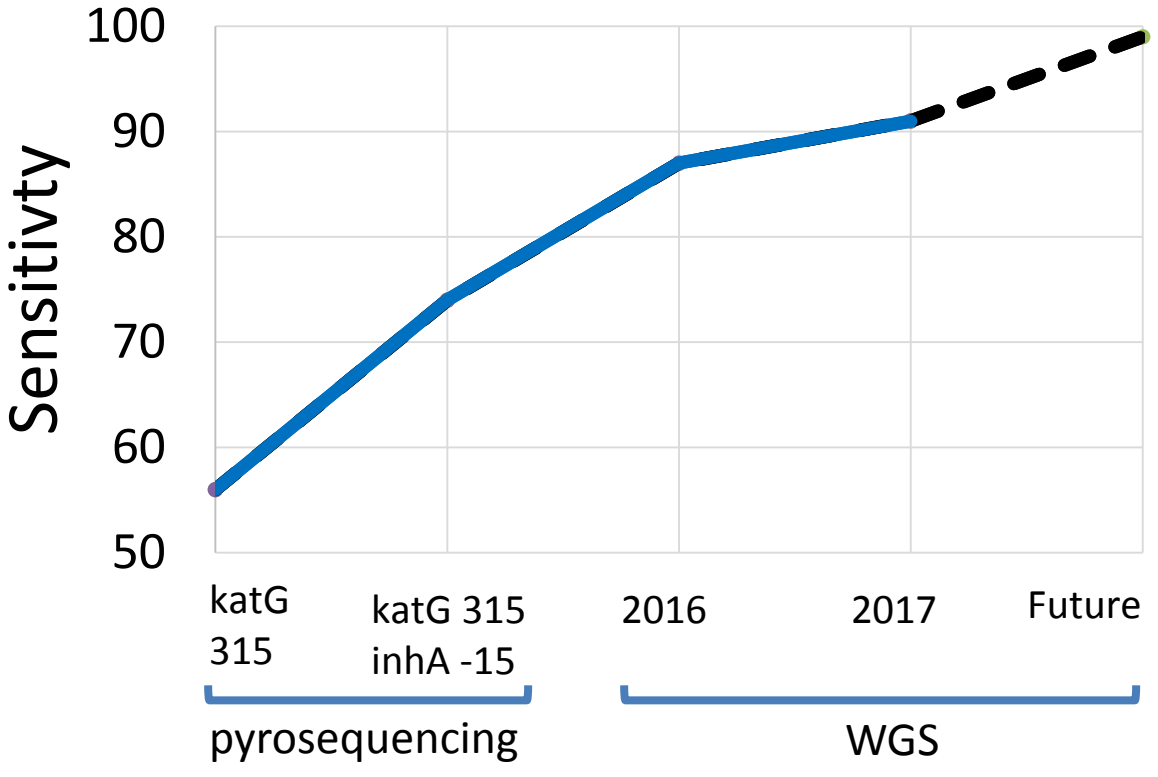
Retrospective Study: Isoniazid comparison

		DST Phenotype	
	Isoniazid	**R**	**S**
WGS Genotype	**R**	55	1[1]
	S	6[2]	32

[1]This SNP is known to be a good but not perfect predictor of INH resistance (14/15 resistant)
[2] Each of the 6 has a different mutation that could potentially account for the missed resistance

Resistance Predictive Value= 98%
Susceptible Predictive value= 84%

Molecular INH Resistance Prediction



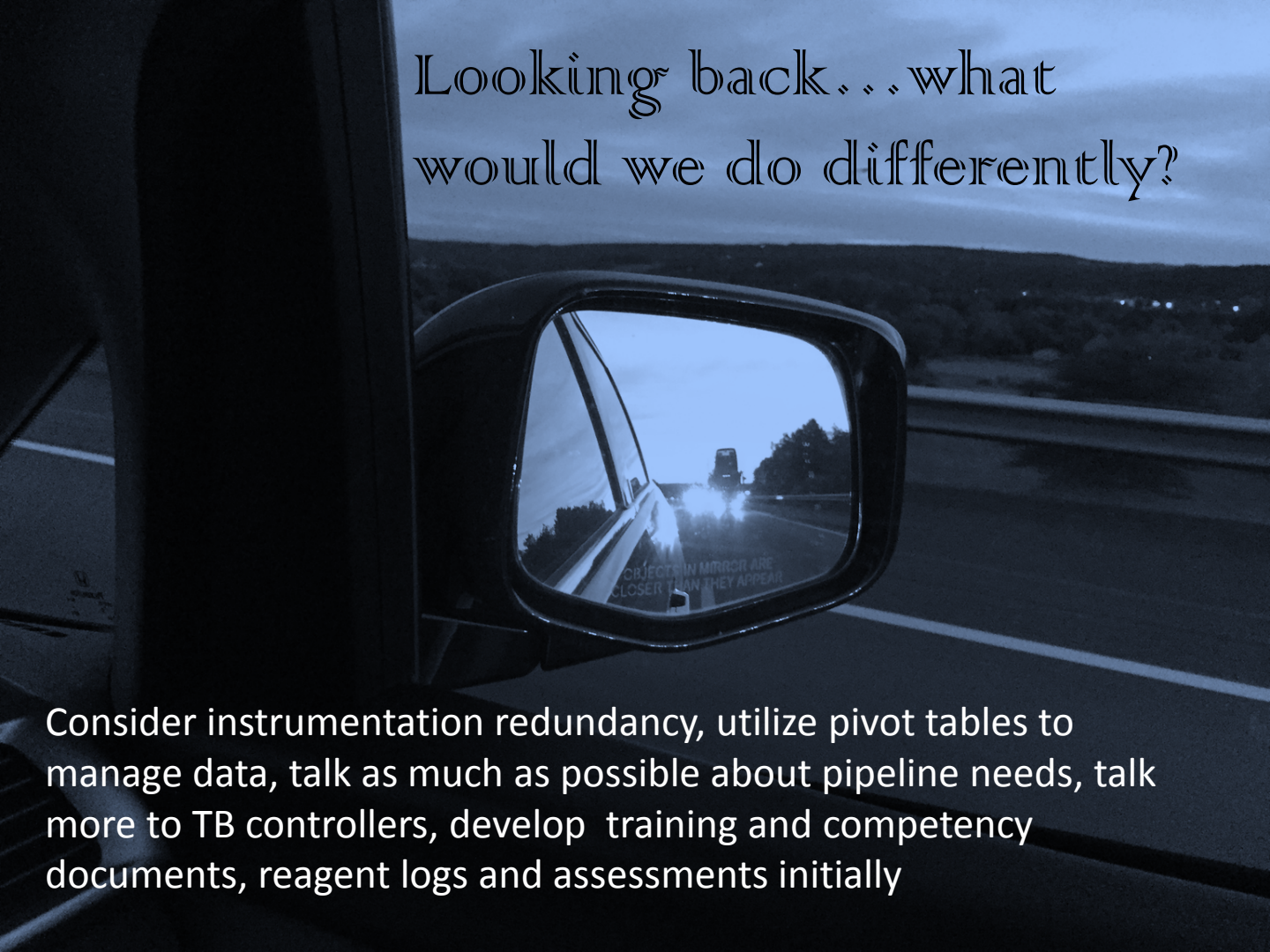
What have we learned in 2+ years…

- Communication with NGS Core and bioinformaticians is critical!
- Discordance almost always is determined to be due to AST
- Not that many surprises, but continual improvement
- TB Control Epidemiologists and Regional NY colleagues love this data!

2015

What have we learned from NYS CLEP and CLIA Surveys?

- This type of testing is new to everyone
- Special internal audit with our QA Officer helpful
- Memos stating assay developers when training documentation doesn't make sense
- Able to utilize CDC Model Performance Evaluation Program (MPEP) for Internal Quality Assurance
- Tracking and QC reagents and Log
- Documenting pipeline updates



Consider instrumentation redundancy, utilize pivot tables to manage data, talk as much as possible about pipeline needs, talk more to TB controllers, develop training and competency documents, reagent logs and assessments initially

Future

- Direct Sputum testing research
- WGS pipeline (2017)
  - –Added drugs AST for MDR strains
  - –WGS targets linezolid, clofazimine, PAS, bedaquiline
  - –thyA stop mutation- PAS resistance
- Third pipeline update (2018)
  - New mutations
  - New mutation category (unclassified)
  - Externally facing pipeline
- Discontinuing DST very conservative approach on strains with no markers for resistance
- Funding
  - NIH R03- Evaluate TB WGS directly on sputum specimens
  - NIH grant award Nanopore MinION TB
  - APHL/CDC RFA to perform TB WGS

How about Antimicrobial Resistance?

- To determine a novel carbapenemase mechanism
- To detect an IMP variant (other than IMP1)
- To assess other resistance genes
- (To determine relatedness)

Why NGS on TB?

- **9000** CRE infections each year

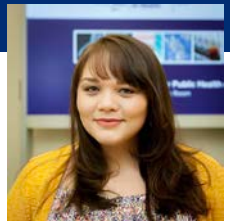
Urgent threat level

Some CRE have become resistant to nearly all available antibiotics- NDM

11% Screening showed hard to treat CRE that can spread easily

Dozens of genes and thousands of variants that can be assessed.

What if there were also random pages included in the pile that needed to stay with the book



Along came an ARLN fellow…

Collect all data
Characterization of isolates
Repeat some initial testing
Build pipeline and validate
Assess WGS
Summarize
Collaborate

Illumina Sequencing Reads

Species ID Check
Multi-locus Sequence Typing

Identification of Antimicrobial Resistance Genes Using:

1. ResFinder
2. ARG-ANNOT
3. NCBI AR Database

Acknowledgements

Core TB WGS Team

Vincent Escuyer Kim Musser
Tanya Halse Joseph Shea
Pascal Lapierre

ARLN Team

Elizabeth Nazarian, Kara Mitchell,
Emily Snavely

MYCOBACTERIOLOGY LAB

MOLECULAR BACTERIOLOGY LAB

APPLIED GENOMIC TECHNOLOGIES CORE

BIOINFORMATICS CORE

Mike Palumbo Erica Lasek-Nesselquist,
Wolfgang Haas

CLIMS

Colleen Walsh, Alok Mehta

Wadsworth CLEP Reviewers, CLEP staff

NYC PHL

Jennifer Rakeman, John Kornblum, Jamie Lemon

Wadsworth Center

Jill Taylor, Victoria Derbyshire, Ron Limberger

CDC

Angela Starks, Jamie Posey

APHL

Anne Gaynor

NYS TB Control, NYC TB Control

Wadsworth Center
NEW YORK STATE DEPARTMENT OF HEALTH

CDC

APHL

Wadsworth Center, NYSDOH
Public Health Genomics Initiative

Establishment of Mycobacterium tuberculosis complex WGS Reference Centers APHL/CDC

NIH National Institutes of Health
Office of Extramural Research

National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention

R03 NIH- Use of whole genome sequencing for tuberculosis diagnostics
*New Funding NIH MinION TB project