ReSeqTB

Relational Sequencing TB Data Platform

A Standardized System for Grading Mutations in *Mycobacterium tuberculosis* for Association with Drug Resistance

PAOLO MIOTTO

CPTR 2017 Workshop, March 20 – 23

The Need

ReSeqTB
Relational Sequencing TB Data Platform

"A lack of user-friendly **data analysis and interpretation tools** has been frequently cited as a major barrier to routine use of WGS techniques."

"Our experience with WGS in the clinical context of drug-resistant TB highlights **the urgent need for an internationally recognized database for standardized genotype-phenotype correlation interpretation** that can be added to and interrogated by clinicians and microbiologists who will be increasingly accessing WGS platforms at the local hospital level."

Wyres KL, et al. Pathogens 2014; 3(2):437-58.
Witney AA, et al. J Clin Microbiol 2015; 53(5):1473-83.

The ReSeqTB Solution: A Standardized System for Grading Mutations

ReSeqTB
Relational Sequencing TB Data Platform

- Driven by Experts and the community
- Grading criteria for confidence binning of mutations
  - Frequency of a mutation in phenotypically resistant or sensitive strain
  - Homoplasy – similar trait acquired by isolates from different lineages
  - Associated MIC values and the degree of change versus a sensitive strain
  - Functional genetics associated with the mutation
- Every SNP will be associated with a "score" or a "grading index" according to the information available
- Grading drug resistance associated mutations
  - 5 bins based upon the strength of associated data for individual mutations
  - Bins are defined using a statistical approach and then revised according to the expert panels confidence of association with resistance of a mutation

Statistical Approach and Thresholds

ReSeqTB
Relational Sequencing TB Data Platform

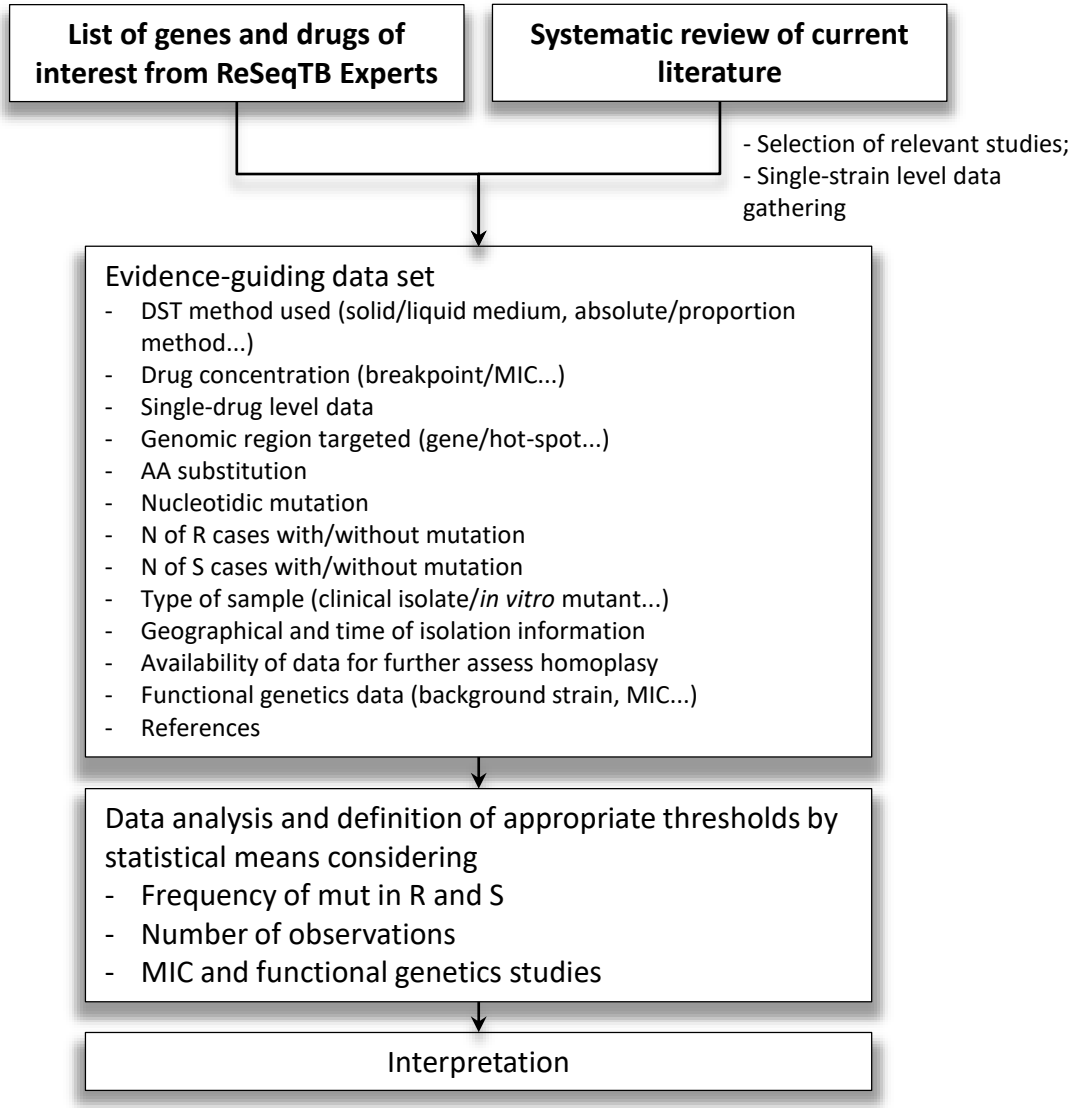
Observed frequency of a mutation found in phenotypically resistant of sensitive strain

- Literature data were used to calculate the frequency of each mutation in resistant (DR) and susceptible (DS) MTBC isolates
- Likelihood ratio (LR) and odd ratio (OR) were used; p-values and 95% confidence intervals associated with LR and OR have been also considered
- Thresholds used most commonly in evidence-based medicine have been adapted to grade *M. tuberculosis* mutations:

LR+ ∨ OR *p*- value	value	Interpretation	Symbol
<0.05	**≥ 10**	**High Confidence for association with resistance** – strong association of the mutation with phenotypic drug resistanc; sufficient evidence that the mutation confers or is strongly associated with drug resistance	
<0.05	**5 ≤ ... < 10**	**Moderate Confidence for association with resistance** – moderate association of the mutation with phenotypic drug resistanc; additional data desirable for improved evidence that the mutation confers or is strongly associated with drug resistance	
<0.05	**1 ≤ ... < 5**	**Minimal Confidence for association with resistance** – weak association of the mutation with phenotypic drug resistanc; inconclusive evidence that the mutation confers or is strongly associated with drug resistance. Substantial additional data required	
<0.05	**< 1**	**No association with resistance** – No evidence of association between the mutation and drug resistance	
≥0.05	-	**Indeterminate** – no statistically significant threshold reached; additional data required	Indeter

Implementation – Systematic review of literature data

ReSeqTB
Relational Sequencing TB Data Platform



Miotto P, et al. Manuscript in preparation

Implementation – Systematic review of literature data

ReSeqTB
Relational Sequencing TB Data Platform

- 11 drugs, 16 genomic loci
- 1748 unique papers screened
- Data: sequencing AND phenotypic DST
- Data from up to 43 countries and up to 13,424 isolates

Drug	Collected data				Number of studies	
	loci of interest	Total # of isolates	Isolation time frame	Represented countries	Screened	Included
Rifampicin (R)	*rpoB*	13424	1999-2014	37	459	95
Isoniazid (H)	*katG*	11847	1992-2014	42	650	127
	inhA-mabA	9407				
	furA	361				
	mshA	288				
Ethionamide & prothionamide (ETO/PTO)	*inhA-mabA*	346				
	ethA	181				
	mshA	117				
Ofloxacin (OFX)	*gyrA*	5911	1991-2013	36	243	75
	gyrB	3078				
Moxifloxacin (MFX)	*gyrA*	1019				
	gyrB	735				
Levofloxacin (LFX)	*gyrA*	449				
	gyrB	218				
Pyrazinamide (Z)	*pncA*	4949	1990-2014	36	378	81
Streptomycin (S)	*rpsL*	3263	1985-2013	43	423	104
	tap	0				
	rrs	2598				
	whiB7	0				
	gidB	812				
Amikacin (AM)	*rrs*	2105				
Capreomycin (CM)	*rrs*	2533				
	tlyA	1854				
Kanamycin (KM)	*rrs*	1727				
	tap	56				
	eis	2029				
	whiB7	56				

Miotto P, et al. Manuscript in preparation

Implementation – Results

ReSeqTB
Relational Sequencing TB Data Platform

Drug (phenotypic testing)		Gene	High confidence mutations	Moderate confidence mutations	Minimal confidence mutations	No association with resistance
First-line	R	*rpoB*	D516A, D516F, D516G, D516G+L533P, D516ins, D516N, D516V, D626E, Del N518, F505V+D516Y, F514dupl, H526C, H526D, H526F, H526G, H526L, H526R, H526Y, M515I+D516Y, Q513-F514ins, Q513H+L533P, S512T, Q513K, Q513L, Q513P, S522Q, S531F, S531L, S531Q, S531W	D516Y, H526P, L533P, S522L	H526N, I572F, L511P	
	H	*inhA-mabA*		c-15t		g-47c, t-80g, T4I, g-102a
		katG	S315I, S315N, S315T, Pooled frameshifts and premature Stop codons			A110V, L499M, R463L
		furA				L68F
		mshA				N111S, A187V
Second-line (group A)	MFX	*gyrA*	A90V, D94A, D94G, D94N, D94Y, G88C, S91P			E21Q, G247S, G668D, S95T, V712L
	OFX/LFX	*gyrA*	A90V, D94A, D94G, D94H, D94N, D94Y, G88A, G88C, S91P	D89N		E21Q, G247S, G668D, S95T, T80A, V712L
		gyrB	A504V, E459K			
Second-line (group B)	AM	*rrs*	a1401g, g1484t			
	KM	*eis*	c-14t, g-10a		c-12t, g-37t	a1338c
		rrs	a1401g, c1402t, g1484t			
	CM	*rrs*	a1401g, c1402t, g1484t			c517t
		tlyA	N236K, Pooled frameshifts and premature Stop codons			D149H
	S	*rpsL*	K43G, K43R, K43T, K88Q, K88R, T40I			
		rrs	a514c, a514t, c462t, c513t, c517t			
		gidB				E92D, L16R, V110G, Pooled frameshifts and premature Stop codons

Miotto P, et al. Manuscript in preparation

Implementation – Results

ReSeqTB
Relational Sequencing TB Data Platform

Drug (phenotypic testing)		Gene	High confidence mutations	Moderate confidence mutations	Minimal confidence mutations	No association with resistance
Second-line (group C)	ETO/PTO	*inhA*	c-15t+I194T, c-15t+S49A	c-15t		
		ethA				Q347Stop
Second-line (group D)	Z	*pncA*	a-11g, A134V, A3E, A46V, C138Y, C14R, C72R, D12A, D12N, D49G, D49N, D63G, D8E, D8G, D8N, F94L, F94S, G108R, G132A, G132D, G132S, G162D, G17D, G24D, G97C, G97D, G97S, H137P, H51Q,H51R, H57D, H57P, H57R, H57Y, H71D, H71Q, H71Y, H82R, I6T, indel - R148ins (inframe), K96N, K96R, L116P, L116R, L120P, L151S, L159P, L172P, L19P, L4S, L85P, L85R, M175T, M175V, P54S, P62L, P62Q, Q10P, Q141P, R123P, S104R, S59P, S66P, S67P, t-12c, T135N, T135P, T142A, T142K, T142M, T160P, T168P, T76P, t-7c, V125F, V125G, V128G, V139G, V139L, V155G, V180F, V180G, V7G, W68C,W68R, Y103H, Y34D, Pooled frameshifts and premature Stop codons	A171E, K96E, K96T, M175I, P54L, Q10R, W68G	D12G, F58L, H71R, I133T, V139A	I31T, I6L, indel - c-125del, K48T, L35R, T114M, T47A

- 389 mutations (including frameshifts and premature stop codons) had minimal to high confidence values with regards to predicting drug resistance in MTBC

Miotto P, et al. Manuscript in preparation

Advantages of the Grading System

ReSeqTB
Relational Sequencing TB Data Platform

Mutations not associated with DR

Drug	Gene	Mutation
H	*katG*	A110V, L499M, R463L
	inhA-mabA	g-102a, g-47c, t-80g, T4I
	mshA	A187V, N111S
MFX	*gyrA*	E21Q, G247S, G668D, S95T, V712L
OFX, LFX	*gyrA*	E21Q, G247S, G668D, S95T, T80A, V712L
KM	*rrs*	a1338c
CM	*rrs*	c517t
	tlyA	D149H
S	*rrs*	a1401g
	gidB	E92D, L16R, V110G, frameshifts and premature Stop codons
Z	*pncA*	I31T, I6L, c-125del, K48T, L35R, T114M, T47A
ETO/PTO	*ethA*	Q347Stop

Identification of potential phenotypic testing reproducibility issues

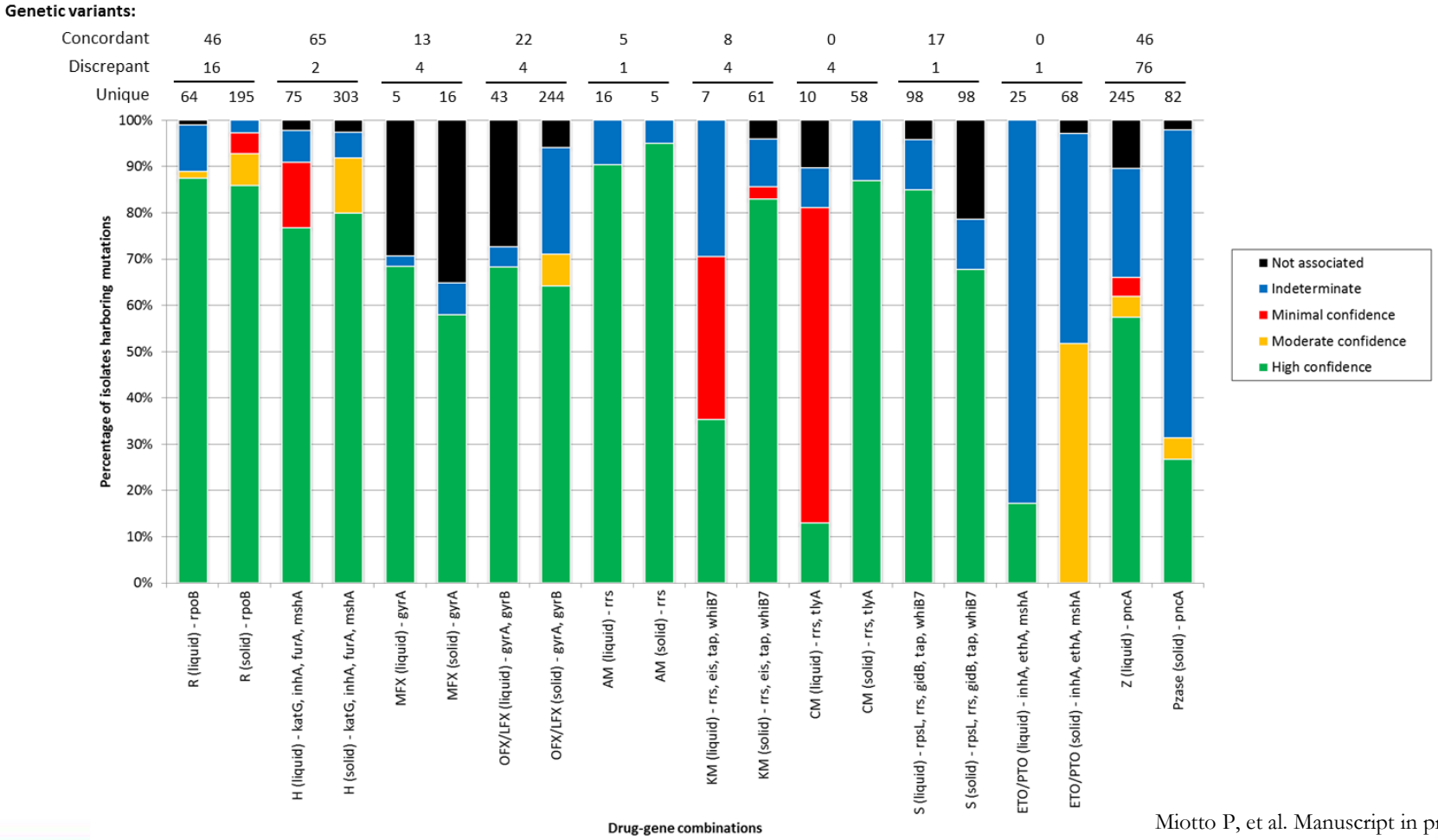
Drug	Gene	Mutation	Found in DR (TP)	Found in DS (FP)	Sensitivity		Specificity		Accuracy		Confidence value
					%	95% CIs	%	95% CIs	%	95% CIs	
R	*rpoB*	S531L	4416	20	53.2	52.2-54.3	99.6	99.4-99.8	71.0	70.2-71.7	
H	*katG*	S315T	4130	30	53.6	52.5-54.7	96.2	94.6-97.4	57.6	56.5-58.6	
OFX/OFX	*gyrA*	A90V	397	4	10.9	9.9-11.9	99.9	99.7-100.0	55.6	54.4-56.7	
AM	*rrs*	a1401g	600	5	74.2	71.0-77.2	99.6	99.1-99.9	89.8	88.5-91.1	
S	*rpsL*	K43R	358	6	62.9	58.8-66.9	98.1	95.8-99.3	75.3	72.3-78.1	
...	...	...	...	...	...	...	...	...	...	...	...

Miotto P, et al. Manuscript in preparation

Advantages of the Grading System

ReSeqTB
Relational Sequencing TB Data Platform

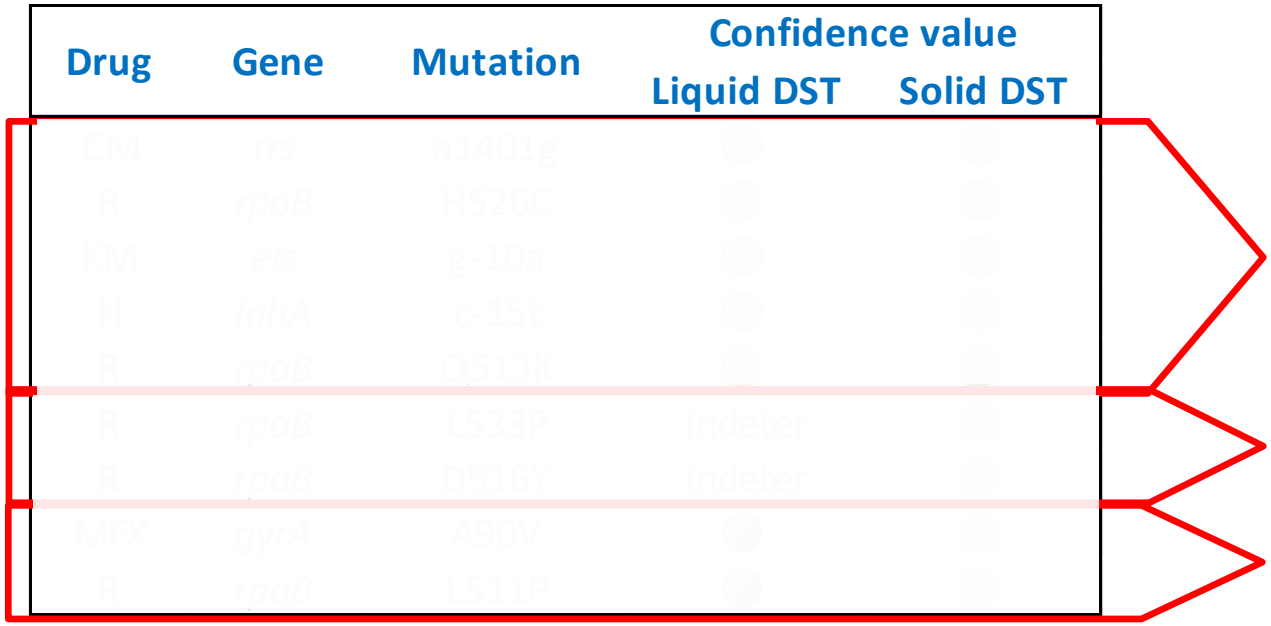
- Specific mutations showed different confidence levels depending upon the medium considered for phenotypic DST
- The number of these discrepancies varied from just one variant for drugs such as AM, S, and ETO/PTO to 76 for Z



Miotto P, et al. Manuscript in preparation

Advantages of the Grading System

ReSeqTB
Relational Sequencing TB Data Platform



Cases testing differently on different media

MIC close to liquid testing critical concetration

Cases missed by liquid phenotypic testing

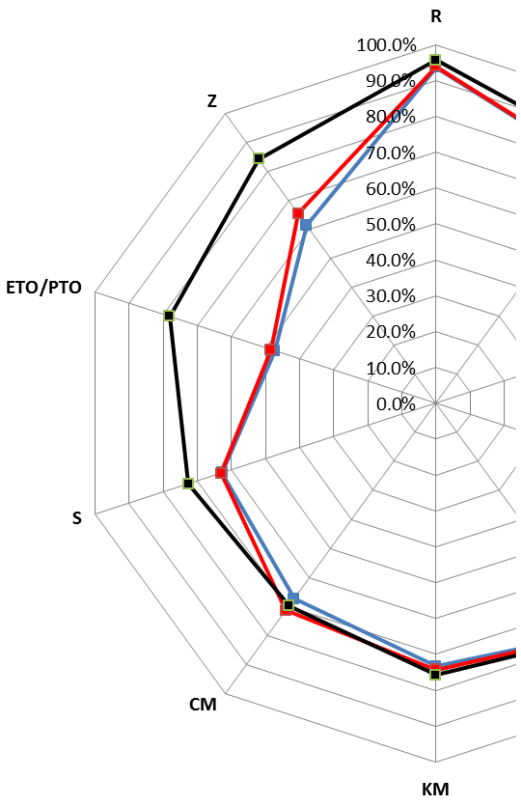
MOX liquid CC: 2 mg/L

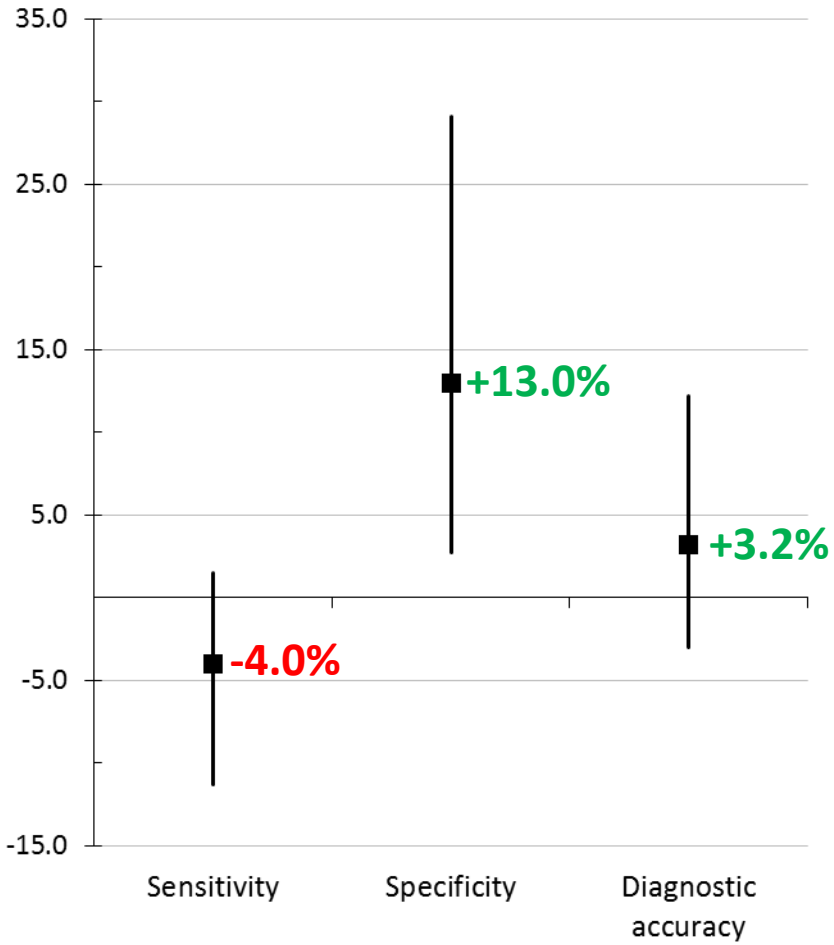
Miotto P, et al. Manuscript in preparation

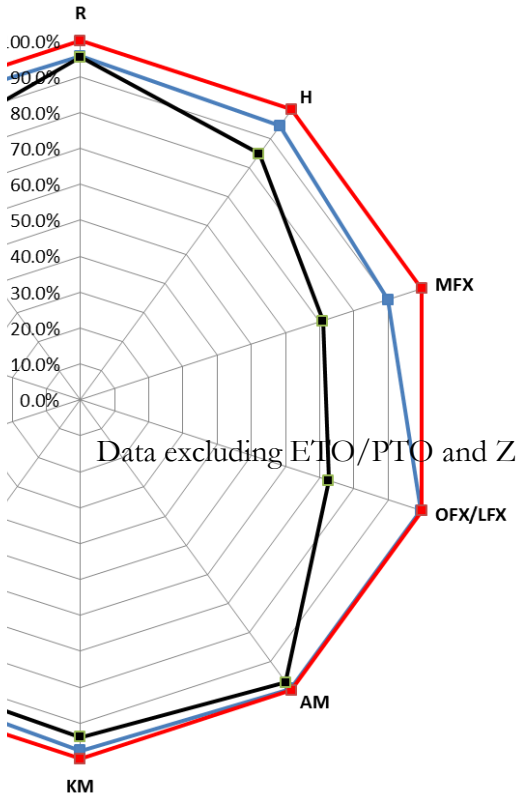
Implications of the Grading System

ReSeqTB
Relational Sequencing TB Data Platform

Sensitivity







Data excluding ETO/PTO and Z

- The use of graded mutations allows for improvement in specificity with a relatively small decrease in sensitivity
- Reduction of Very Major Errors (DR reported as DS)

Miotto P, et al. Manuscript in preparation

Summary

ReSeqTB
Relational Sequencing TB Data Platform

- A total of 389 high, moderate, and minimal confidence mutations have been determined

- Advantages of the grading system:
  - suitable for automated classification of mutations
  - LR is a universal measure of association which is not affected by local prevalence
  - unlike sensitivities and specificities, LRs do not lead to an exaggeration of the benefits of a test or strength of an association

- The validation process is driven by statistics based on frequency of mutations, this will be improved by integrating
  - Homoplasy data
  - MIC data
  - functional genetic studies

- The validation process will need to be evaluated periodically and refined as tools and our understanding increase

Acknowledgements

Timothy Rodwell (FIND, Switzerland)

David Dolinger (FIND, Switzerland)

Leonid Chindelevitch (Simon Fraser University, BC, Canada)

Naomi Hillery (UCSD, CA, USA)

Elisa Tagliani (San Raffaele Scientific Institute, Italy)

Belay Tessema (FIND, Switzerland)

ReSeqTB Input and Output expert panels and the ReSeqTB team at the CPTR

S Niemann, J Fleming, R Warren, D Crook, C Denkinger, J Posey, S Gagneux, S Hoffner, C Rodrigues, CU Köser, I Comas, D Engelthaler, M Murray, D Alland, L Rigouts, C Lange, K Dheda, R Hasan, U Devi, R McNerney, M Ezewudo, M Schito, DM Cirillo