The interpretation algorithm for TBProfiler_tNGS_PHB pipeline output was developed by Microbial Diseases Laboratory (MDL) of California Department of Public Health for integration in the routine TB DST workflow and may not be suitable for all laboratories depending on the availability of confirmatory phenotypic testing and expertise in genetics of TB resistance. The interpretation algorithm is subject to change. For any inquiries, please contact MDL Mycobacterial, Mycotic, and Parasitic Diseases Section, Dr. Varvara Kozyreva (varvara.kozyreva@cdph.ca.gov) and Dr. Matthew Sylvester (matthew.sylvester@cdph.ca.gov)

Principles of Analysis

WGS- and tNGS-based DST assays utilize the same pipeline with some scripts and analysis parameters specific to each of the assays. Mutation interpretation is the same for both WGS-DST and tNGS DST.

WGS:

MTBC WGS-DST bioinformatics analysis is done using in-house developed TB DST WGS bioinformatics workflow (TheiaProk_Illumina_PE_PHB v.1.3.0, [PHB]). This document provides a detailed description of the pipelines and databases included into the workflow, as well as parameters of analysis.

As an overall principle of bioinformatics analysis performed by TheiaProk_Illumina_PE_PHB, the thousands of sequencing reads that were generated across MTBC genome are mapped to a reference sequence *Mycobacterium tuberculosis* H37Rv NC_000962.3 and analyzed for genetic differences (i.e., mutations) [the current version of PHB v1.3.0 was validated with the following components: gambit v. 1.0.0, TB Profiler v. 4.4.2, and tbp-parser:1.2.2]. The mutations detected in the specific genomic loci that are known to be associated with drug resistance in MTBC are extracted and filtered based on information about mutations with known effects by querying the database of mutations [TBDB v.2023-03-26]. The following fixed (not exposed for modification by user) parameters are implemented in the validated version of the WGS workflow:

tbprofiler_min_af_pred = 0.1
tbprofiler_cov_frac_threshold = 0
tbprofiler_min_af = 0.1
tbprofiler_min_depth = 10
tbp_parser_min_depth = 10
tbp_parser_coverage_threshold = 100

tNGS:

MTBC tNGS-DST bioinformatics analysis is done using in-house developed TB DST tNGS bioinformatics workflow (TBProfiler_tNGS_PHB v.1.0.0). This document provides a detailed description of the pipelines and databases included in the workflow, as well as parameters of analysis.

As an overall principle of bioinformatics analysis performed by TBProfiler_tNGS_PHB, the sequencing reads that were generated across targeted regions of MTBC genome are mapped to a reference sequence Mycobacterium tuberculosis H37Rv NC_000962.3 and analyzed for genetic differences (i.e., mutations). [TBProfiler_tNGS_PHB v1.0.0 was validated with the following components: TB Profiler v. 4.4.2 and tbp-parser:1.4.4.5]. The mutations detected in the specific genomic loci that are known to be associated with drug resistance in MTBC are extracted and filtered based on information about mutations with known effects by querying the database of mutations [TBDB v.2023-03-26].

The following input parameters have been validated with TBProfiler_tNGS_PHB v1.0.0 (only non-default values are listed):

tbprofiler_tngs.tbprofiler.min_af_pred	0.05
tbprofiler_tngs.tbp_parser.sequencing_method	tNGS
tbprofiler_tngs.tbprofiler.min_af	0.05
tbprofiler_tngs.tbp_parser.docker	us-docker.pkg.dev/general-theiagen/theiagen/tbp-parser:1.4.2"
tbprofiler_tngs.tbp_parser.rpob449_frequency	0.9
tbprofiler_tngs.tbp_parser.rrs_read_support	20
tbprofiler_tngs.bases_to_crop	0
tbprofiler_tngs.tbp_parser.rrl_frequency	0.9
tbprofiler_tngs.tbprofiler.min_depth	5
tbprofiler_tngs.tbp_parser.rrs_frequency	0.95
tbprofiler_tngs.tbp_parser.etha237_frequency	0.9
tbprofiler_tngs.tbp_parser.rrl_read_support	35

The following fixed (not exposed for modification by user) parameters are implemented in the validated version of the workflow:

tbprofiler_cov_frac_threshold = 0
tbp_parser_min_depth = 10
tbp_parser_coverage_threshold = 100

For both WGS and tNGS, data about the effects of mutations on drug resistance in *Mycobacterium tuberculosis* complex was obtained from WHO, and select peer-reviewed publications based on extensive correlation with phenotypic DST data. If the detected mutation is not found in the database, additional "expert rules" are applied for interpretation of the mutations within genomic regions for which a substantial amount of evidence was accumulated, suggesting likely resistance in case of presence of certain types of mutations anywhere within those regions. As per WHO, the mechanisms by which mutations in such genes confer resistance are well understood, and no epistatic interaction has been observed that could render an isolate with such mutation susceptible. Otherwise, in the absence of strong evidence of association of the detected mutation with either a resistant (R) or susceptible (S) phenotype, mutations are either reported as possessing uncertain significance (U) or not reported at all, depending on the particular gene and level of the confidence in available phenotypic data as per WHO.

In general, if there is an entry in the TBDB database for a mutation, that takes precedence and is applied first. TBDB database includes the default set as well as tbdb.other_annotations.csv. TBDB watchlist is active.

The described below interpretation logic and expert rules are implemented in the tbp-parser tool available on GitHub.

WGS and tNGS -DST interpretation summary

	Individual mutation reporting		Drug interpretation reporting
Mutation Interpretation	**Format of individual mutations reported per gene locus**	**Mutation is listed in clinical report (Yes/No)**	**Overall value based on the highest severity mutation within all targets associated with given drug**
R-mutations for all targets, except for *rpoB* (see below)	c.2066C>T (p.Ala689Val)	Yes	Mutation(s) associated with resistance to XXX detected
R-mutations in *rpoB* gene: Low-level RIF R mutations	c.2066C>T (p.Ala689Val)	Yes	Predicted low-level resistance to rifampin. May test susceptible by phenotypic methods
R-mutations in *rpoB* gene: Other RIF R mutations	c.2066C>T (p.Ala689Val)	Yes	Predicted resistance to rifampin
U	c.2066C>T (p.Ala689Val)	Yes	The detected mutation(s) have uncertain significance. Resistance to XXX cannot be ruled out
S-mutations for all targets, except for *rpoB* (see below)	No high confidence mutations detected	No	No mutations associated with resistance to XXX detected
S-mutations in *rpoB* gene: Synonymous mutation present within *rpoB* RRDR (codons 426-452)	c.2066C>T (p.Ala689Ala) [synonymous] [1]	Yes	Predicted susceptibility to rifampin. The detected synonymous mutation(s) do not confer resistance [2] Additionally displayed in comments: The detected in *rpoB* synonymous mutation may result in false-resistance in PCR-based assays targeting the *rpoB* RRDR
S-mutations in *rpoB* gene: Other RIF S mutations (outside of RRDR)	No high confidence mutations detected	No	Predicted susceptibility to rifampin [2]
WT	No mutations detected	N/A	No mutations associated with resistance to XXX detected
Insufficient Coverage in the gene locus[3] AND successfully sequenced areas of the same gene OR other genes associated with given drug do NOT contain mutations interpreted as R	No sequence[4]	N/A	Not all targets could be sequenced; resistance to XXX cannot be ruled out[5]

1. Displayed in addition to any “R” or “U” mutations detected in RRDR.
2. If other *rpoB* mutations have interpretation “U” or “R”, then report drug predicted susceptibility based on those mutations’ interpretation.
3. Deletions and areas of poor sequencing coverage are differentiated by bioinformatics pipeline.
4. If portion of the gene has poor coverage but “R” mutation has been detected within different area of the same gene that passes QC threshold, report corresponding drug interpretation based on present “R” mutation.
5. If other genes associated with the same drug have mutations with interpretation “R”, then report drug interpretation based on mutations in those targets.

Genomic Regions of Interest- WGS

Start nt position	Stop nt position	Gene name	Associated with resistance to
5040	7467	*gyrB*	moxifloxacin, levofloxacin
7102	10018	*gyrA*	moxifloxacin, levofloxacin
759607	763525	*rpoB*	rifampicin
775386	778680	*mmpL5*	clofazimine, bedaquiline
778277	779105	*mmpS5*	clofazimine, bedaquiline
778790	779687	*Rv0678*	clofazimine, bedaquiline
800609	801662	*rplC*	linezolid
1460845	1461490	*atpE*	bedaquiline
1471646	1473582	*rrs*	kanamycin, capreomycin, amikacin
1473458	1476995	*rrl*	linezolid
1673148	1674383	*fabG1*	ethionamide, isoniazid
1673848	1675211	*inhA*	ethionamide, isoniazid
1917740	1918946	*tlyA*	capreomycin
2153689	2156570	*katG*	isoniazid
2288481	2290323	*pncA*	pyrazinamide
2713924	2715586	*eis*	kanamycin, amikacin
2859100	2860618	*pepQ*	clofazimine, bedaquiline
4243004	4246717	*embA*	ethambutol
4246314	4250010	*embB*	ethambutol
4325804	4330174	*ethA*	ethionamide

Genomic Regions of Interest- tNGS

Range	Broad Reportable Range		Essential for Resistance Range			
gene	start nt position*	end nt position*	start nt position*	end nt position*	All R and R-Interim mutations[3] captured by ERR	FL region under expert rule is covered
eis	2715171	2715528	2715171	2715421	Yes	N/A
embB	4247376	4248065	4247376	4248065	Yes	N/A
ethA	4325951	4327510	4326003	4327485	Yes	Yes
fabG1	1673321	1673755	1673353	1673755	No[1]	N/A
gyrA	7377	7754	7383	7755	Yes	Yes
gyrB	6298	6943	6298	6943	Yes	Yes
inhA	1674287	1674880	1674287	1674880	Yes	N/A
katG	2153404	2156137	2153888	2156114	Yes	Yes
pncA	2288672	2289301	2288680	2289301	Yes	Yes
rplC	801108	801483	801108	801462	Yes	No[2]
rpoB_1	760957	761355	760957	761355	Yes	Yes
rpoB_2	760280	760812	760280	760812	Yes	N/A
rrl	1475923	1476625	1475923	1476619	Yes	No
rrs	1471850	1473945	1472182	1473382	Yes	Yes
Rv0678	778990	779487	778990	779487	Yes	No[2]
tlyA	1917811	1918750	1917933	1918746	Yes	N/A

Footnotes: “Broad Reportable Range” (BRR)- is the reportable genomic range of tNGS-DST assay within which we have demonstrated ability to consistently obtain sufficient coverage depth and accurate mutation detection. In some cases, however, it is not possible to obtain a complete coverage of loci within BRR, hence, the regions encompassing high-confidence resistance mutations or regions affected by expert rules were established for each locus and referred to as regions within the “Essential for Resistance Range” (ERR). In cases when less than 100% of BRR covered but 100% of ERR was successfully sequenced, it is acceptable for SME to convey this information to the submitter notifying them that even though we cannot exclude a possibility of resistance-conferring mutations occurring outside of ERR, the presence of R mutations outside of the area that have been successfully sequenced is less likely.

*Position coordinates listed in relation to the *M. tuberculosis* H37Rv NC_000962.3

1. Deeplex design misses one R-I mutation in *fabG1* position 1674048, but all others are captured.

2. FL protein-encoding region for *rplC* and *Rv0678* are covered, however, the promoter regions are not.
3. As per WHO v.2: Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance, Second edition, WHO, 2023. https://www.who.int/publications/i/item/9789240082410

MDL WGS-DST interpretation logic

Last updated: 05/31/2024
TBP Parser v.1.4.4.5

In general, if there is an entry in the TBDB database for a mutation, that takes precedence and is applied first. TBDB database includes the default set as well as tbdb.other_annotations.csv. TBDB watchlist is active.

Below in the sections 1-4, please see the detailed description of parsing that is performed to generate interpretations for detected mutations, also referred to as "Laboratorian report" and contains broader information than is reported as a part of clinical report (see section 5).

1. For **genes *mmpR5 (Rv0678), atpE, pepQ, mmpL5, mmpS5, rrl, rplC* (*genes related to new drugs based on CDC expert rules*)**:

1.1. IF mutation is in either of *mmpR5 (Rv0678*), *atpE, pepQ, mmpL5, mmpS5, rrl, rplC*
AND who_confidence has value for the corresponding drug,
THEN keep that value as in "confidence" column **AND** assign the following "Looker interpretation" & "MDL Interpretation" based on that confidence value:

Mutation classification (confidence field)	**Looker interpretation**	**MDL Interpretation**
WHO: associated with R [1]	R	R
WHO: associated with R- interim [2]	R-Interim	R
WHO: uncertain [3,4]	U	U
WHO: NOT associated w/R [5]	S	S
WHO: NOT associated w/R- interim [6]	S-Interim	S
WHO: synonymous [7]	S	S

[1] Exact value in WHO & TBDB "1) Assoc w R"
[2] Exact value in WHO & TBDB "2) Assoc w R – Interim"
[3] Exact value in WHO & TBDB "3) Uncertain significance"
[4] WHO also has value "NA" that is not found in TBDB
[5] Exact value in WHO "5) Not assoc w R". Not in TBDB
[6] Exact value in WHO & TBDB "4) Not assoc w R – Interim"
[7] Exact value in WHO "Synonymous". Not in TBDB

1.2. IF mutation is in either of *mmpR5 (Rv0678), atpE, pepQ, mmpL5, mmpS5, rrl, rplC* **AND no** who_confidence available for the corresponding drug **THEN** assign confidence= "No WHO annotation" **AND** assign the "Looker interpretation" & "MDL Interpretation" value based on table below:

Genes	Mutation location	Mutation type	Looker interpretation	MDL Interpretation
mmpR5 (Rv0678), atpE, pepQ, rplC	Target Promoter* (see coordinates below)	any	U	U
	Promoter (effect type= upstream_gene_variant, but not within the target promoter region)	any	U	S
	ORF (everything else)	NONsyn	U	U
		syn	S	S
mmpL5	Promoter (effect type= upstream_gene_variant)	any	U	S
	ORF (everything else)	NONsyn	U	U
		syn	S	S
mmpS5	Promoter (effect type= upstream_gene_variant)	any	U	S
	ORF (everything else)	NONsyn	U	U
		syn	S	S
rrl	nt positions 2003-2367 and 2449-3056 in rRNA	any	U	U
	outside of nt positions 2003-2367 and 2449-3056 in rRNA	any	U	S

* Promoter regions for different genes are defined here as:

	Promoter region coordinates
rplC	-18 to -1
mmpR5 (Rv0678)	-84 to -1
atpE	-48 to -1
pepQ	-33 to -1

1.3. Only for genes *mmpR5 (Rv0678), mmpL5, mmpS5*- to standardize the output and interpretation of mutations that may be listed in relation to any of those 3 genes:

1.3.1. Output mutations in "alternative_consequences" field to get information about these mutations in relation to the other three genes
Follow rule 1.2 to assign Looker interpretation and MDL interpretation.

2. For **genes *katG, pncA, ethA, gid, rpoB (covered by WHO expert rules):***

2.1. IF mutation is in either of *katG*, *pncA*, *ethA*, *gid, rpoB*,
AND who_confidence has value for the corresponding drug,
THEN keep that value as in "confidence" column **AND** assign the following "Looker interpretation" & "MDL Interpretation" based on that confidence value:

Mutation classification (confidence field)	**Looker interpretation**	**MDL Interpretation**
WHO: associated with R	R	R
WHO: associated with R- interim	R-Interim	R
WHO: uncertain	U	U
WHO: NOT associated w/R	S	S
WHO: NOT associated w/R- interim	S-Interim	S
WHO: synonymous [8]	S	S

2.2. IF mutation found in either of *katG, pncA, ethA, gid, rpoB,*
AND no who_confidence available for the corresponding drug,
THEN apply WHO Expert rule:

2.2.1. IF mutation found in either of *katG, pncA, ethA, gid (*not rpoB*)*:

2.2.1.1. IF mutation represents a loss-of-function (mutation contains "del", "ins", "fs", "delins", "_" or ends with "*")
AND
Those mutations are found in ORF or within first 30 nucleotides upstream start codon
THEN assign confidence= "No WHO annotation" **AND** assign the "Looker interpretation" & "MDL Interpretation" = "**R**" for corresponding mutation.
ELSE assign confidence= "No WHO annotation" **AND** assign the "Looker interpretation" & "MDL Interpretation" value based on table below:

[8] Not the same as synonymous mutations that are not found in WHO and labeled as such.

Mutation classification	Looker interpretation	MDL Interpretation
Detected mutation not in WHO, not covered by expert rule AND **synonymous**	S	S
Detected mutation not in WHO, not covered by expert rule AND **NONsynonymous** [9]	U	U
Detected mutation not in WHO, not covered by expert rule AND **Promoter variants with effect type= upstream_gene_variant**	U	S

2.2.2.IF mutation found in *rpoB*:

2.2.2.1. **IF** mutation within *rpoB* codons 426-452 (even partially, if it's indel) **THEN** assign confidence= "No WHO annotation" **AND** assign the "Looker interpretation" & "MDL Interpretation" value based on table below:

Mutation classification	Looker interpretation	MDL Interpretation
Detected mutation not in WHO AND **synonymous**	S	S
Detected mutation not in WHO AND **NONsynonymous** [10]	R	R

2.2.2.2. **IF** mutation outside of codons 426-452 **THEN** assign confidence= "No WHO annotation" **AND** assign the "Looker interpretation" & "MDL Interpretation" value based on table below:

Mutation classification	Looker interpretation	MDL Interpretation

[9] Including indels in ORF

[10] Including indels in ORF

Detected mutation not in WHO and not covered by expert rule; **synonymous**	S	S
Detected mutation not in WHO and not covered by expert rule; **NONsynonymous** [11]	U	U
Detected mutation not in WHO and not covered by expert rule; **Promoter variants with effect type= upstream_gene_variant**	U	S

3. For genes **other than** *mmpR5 (Rv0678), atpE, pepQ, mmpL5, mmpS5, rrl, rplC, katG, pncA, ethA, gid, rpoB*:

3.1. IF mutation is in gene other than *mmpR5 (Rv0678), atpE, pepQ, mmpL5, mmpS5, rrl, rplC, katG, pncA, ethA, gid, rpoB*
AND who_confidence has value for the corresponding drug,
THEN keep that value as in "confidence" column
AND assign the following "Looker interpretation" & "MDL Interpretation" based on that confidence value:

Mutation classification (confidence field)	**Looker interpretation**	**MDL Interpretation**
WHO: associated with R	R	R
WHO: associated with R- interim	R-Interim	R
WHO: uncertain	U	U
WHO: NOT associated w/R	S	S
WHO: NOT associated w/R- interim	S-Interim	S
WHO: synonymous [12]	S	S

3.2. IF mutation is in gene other than *mmpR5 (Rv0678), atpE, pepQ, mmpL5, mmpS5, rrl, rplC, katG, pncA, ethA, gid, rpoB*
AND no who_confidence available for the corresponding drug

3.2.1. IF Mutation is located within *rrs* gene region,
3.2.1.1. IF Mutations at *rrs* nucleotide positions 1401, 1402, or 1484 (positions in relation to NC_000962.3: 1473246, 1473247, or 1473329)
THEN assign:

[11] Including indels in ORF
[12] Not the same as synonymous mutations that are not found in WHO and labeled as such.

confidence= "No WHO annotation";
"Looker interpretation" = U
"MDL Interpretation" = U

3.2.1.2. **IF** Mutations at *rrs* nucleotide positions OTHER THAN 1401, 1402, or 1484 (positions in relation to NC_000962.3: 1473246, 1473247, or 1473329),
THEN assign:
confidence= "No WHO annotation";
"Looker interpretation" = U
"MDL Interpretation" = S

3.2.2.IF Mutation is located within *gyrA* coding gene (GyrA QRDR expert rule):

3.2.2.1. **AND** mutation is within region codons 88-94 of *gyrA* AND mutation is nonsynonymous
THEN
confidence= "No WHO annotation"
"Looker interpretation" = U
"MDL Interpretation" = U

3.2.3.IF Mutation is located within *gyrB* coding gene (GyrB QRDR expert rule):

3.2.3.1. **AND** mutation is within region codons 446-507 of *gyrB* AND mutation is nonsynonymous
THEN
confidence= "No WHO annotation"
"Looker interpretation" = U
"MDL Interpretation" = U

3.2.4. OTHERWISE (all remaining scenarios not covered above),
THEN assign confidence= "No WHO annotation"
AND
assign the "Looker interpretation" & "MDL Interpretation" value based on table below:

Mutation classification	**Looker interpretation**	**MDL Interpretation**
Detected mutation not in WHO and not covered by expert rule; **synonymous**	S	S
Detected mutation not in WHO and not covered by expert rule; **NONsynonymous** [13]	U	U
Detected mutation not in WHO and not covered by expert rule; **Promoter variants with effect type= upstream_gene_variant**	U	S

[13] Including indels in ORF

4. Reporting of remaining scenarios:

4.1. IF No mutations detected in given gene (WT)
AND
Gene locus passes QC based on coverage report
THEN
add that gene at the bottom of laboratorian report AND assign:
tbprofiler_variant_substitutions = "WT"
n_mutation, aa_mutation = "WT"
confidence, depth, frequency, read_support, rationale = "N/A"
antimicrobial = corresponding drug to which this gene locus confers resistance (may have to list the same locus more than once as "WT" for each of the drugs that it is associated with, since mutations in some loci cause R to more than one drug)
Looker interpretation = "S"
MDL Interpretation = "WT"

4.2. Quality filtering of final reportable results:
QC ranges for determining breadth of coverage were established with the following logic: (1) For each gene, the region is initially established by using the coding region boundaries; (2) then, extend the region 30bp upstream UNLESS that gene has a specific promoter region specified in expert rule 1.2 that extends beyond 30bp upstream.

4.2.1. QC in the position where mutation is detected (see in Laboratorian report)

4.2.1.1. **IF** Detected mutation that is NOT a Deletion failed QC in that position in any of the following QC parameters:
Total Read Depth ("depth" column in Lab report) < 10X
Variant Read Depth ("read_support" column in Lab report) < 10X
Percent Alt Allele ("frequency" column in Lab report) < 10%
[Or alternative min quality thresholds established for specific genes and mutations listed in "Principles of analysis" section of this document.]
THEN
Write "Failed quality in the mutation position" in the *Warning field* for the corresponding mutation.

4.2.1.2. **IF** Detected DELETION has Total Read Depth or/and Variant Read Depth >0 **AND** < 10x [or less than an alternative min total read depth or variant read depth for specific gene/mutation]
THEN
Write "Failed quality in the mutation position" in the *Warning field* for the corresponding mutation.

4.2.1.3. **ELSE** report mutation on LIMS report (includes deletions that have Total Read Depth or/and Variant Read Depth = 0 but passing Percent Alt Allele threshold, since for those the depth of coverage in TB Profiler v. 4.4.2. is not evaluated correctly)

4.2.1.4. Notes:

4.2.1.4.1. Mutations (deletions or other types of mutations) that failed quality in position do not appear on LIMS report and do not affect drug interpretation. E.g., if mutation is S and failed quality in position, it will not affect LIMS drug interpretation, and instead of gene result appearing as "No high confidence mutations detected" it will display

"No mutations", if no other higher severity mutations are found in the gene.

4.2.2. Breadth of coverage throughout the locus: 100% of the locus must be covered with at least 10x; reflected in coverage report

4.2.2.1. **IF** breadth of coverage at 10x is = 100% = PASS, report as is.

4.2.2.2. **IF** breadth of coverage at 10x is < 100% **AND** a deletion is present= PASS, report as is.

4.2.2.3. **IF** breadth of coverage at 10x is < 100% and a deletion is absent= FAIL

AND

4.2.2.3.1. No mutations detected (WT)
THEN
Add "Insufficient coverage for the locus" in the *Warning field* for the corresponding gene
AND
Overwrite the following values for that gene (found at the bottom of laboratorian report):
Looker interpretation= "Insufficient Coverage"
MDL Interpretation= "Insufficient Coverage"
Do not change any other fields from what they normally would be for a WT locus.

OR

4.2.2.3.2. Only "S" or "U" mutations are detected in given gene (i.e., no "R" mutations detected based on MDL interpretation)
THEN
Add "Insufficient coverage for the locus" in the *Warning field* for the corresponding detected S or U mutation, if any (in addition to any other warning from the variation position QC; do not overwrite).
AND
Overwrite the following values:
Looker interpretation= "Insufficient Coverage"
MDL Interpretation= "Insufficient Coverage"
Do not change any other fields from what they normally would be for the detected mutation.

OR

4.2.2.3.3. R mutation is detected in given gene (based on mutation interpretation in "MDL interpretation" field) and **NO** "Failed quality in the mutation position" in the *Warning field*
THEN

Write "Insufficient coverage for the locus" in the *Warning field* (do not overwrite Looker/MDL interpretation columns)

OR

4.2.2.3.4. R mutation detected in given gene (based on mutation interpretation in "MDL" interpretation field) **AND** "Failed quality in the mutation position" in the *Warning field*

THEN

Add "Insufficient coverage for the locus" in the *Warning field* for the corresponding detected mutation (in addition to any other warning from the variation position QC; do not overwrite).

AND

Overwrite the following values:
Looker interpretation= "Insufficient Coverage"
MDL Interpretation= "Insufficient Coverage"
Do not change any other fields from what they normally would be for the detected mutation.

4.3. Additional analysis notes:

4.3.1. Mutations (indels) that start outside of target regions that are a subject of an expert rule, but span into the region of interest, are interpreted as meeting the corresponding rule. E.g. a deletion in rplC occurs over -24 to -17, part of it would be in the promoter region coordinates for that gene (-18 to -1). Or indel starts outside of RRDR but continues into RRDR region. This rule covers indels that completely encompass region of interest.

4.3.2. If sample has less than 100% of the locus covered with at least 10x (failed breadth of coverage), subject matter expert must review the sequence manually to confirm absence of large deletions that may be missed by the pipeline. If a deletion of >=50 bp detected in the gene of interest, SME may report such mutation as "U".

4.3.3. All antimicrobial drugs associated with a gene are reported. This includes WHO annotation and the gene_associated_drugs field. I.e., when there is a mutation in the gene associated with two drugs (e.g. Rv0678 for BDQ and CFZ) and TBDB only has confidence listed for one drug and not another drug, the following logic was implemented:

4.3.3.1. Generate interpretation based on confidence in TBDB for the drug that is listed in TBDB;

4.3.3.2. For the drug that is not listed in the TBDB for that mutation- follow applicable expert rule.

5. LIMS report parsing:

5.1. Below is description of how a separate output file is generated for ingest into LIMS[14] by parsing results of described above interpretation output (aka "Laboratorian report") for resistance reporting (5.2-5.3), parsing of TB Profiler output for species ID (5.4).

5.2. Outputting ***drug interpretations*** in LIMS report (e.g. M_DST_B01_INH): an overall value based on the highest-severity mutation (from high to low: R > U > S > WT) is generated. Assign text value for corresponding drug according to the table below.

<table>
<tr><th colspan="3">Information coming from Laboratorian report</th><th>In LIMS output of TB Profiler</th></tr>
<tr><td colspan="3">MDL Interpretation (overall value based on the highest severity mutation within all targets associated with given drug)</td><td>Drug (e.g. M_DST_B01_INH)</td></tr>
<tr><td rowspan="3">R</td><td rowspan="2">R mutations in rpoB gene</td><td>Following mutations in rpoB gene:

Leu430Pro
Asp435Tyr
His445Asn
His445Ser His445Leu
His445Cys
Leu452Pro
Ile491Phe
AND
No other R mutations in rpoB gene</td><td>Type "Predicted low-level resistance to rifampin. May test susceptible by phenotypic methods"</td></tr>
<tr><td>All other R mutations in rpoB gene</td><td>Type "Predicted resistance to rifampin"</td></tr>
<tr><td colspan="2">R- for all other cases</td><td>Type "Mutation(s) associated with resistance to XXX detected"</td></tr>
<tr><td>U</td><td colspan="2"></td><td>Type "The detected mutation(s) have uncertain significance. Resistance to XXX cannot be ruled out"</td></tr>
<tr><td rowspan="2">S</td><td colspan="2">S for all drugs, except RIF (see below)</td><td>Type "No mutations associated with resistance to XXX detected"</td></tr>
<tr><td>S interpretation for RIF</td><td>Synonymous mutation present within rpoB codons 426-452 (RRDR)
AND
No other mutations interpreted as R or U are detected in rpoB (anywhere in the gene)

P.S.: if synonymous mutations outside of rpoB codons 426-452 follow the rule for "S" above.</td><td>Type "Predicted susceptibility to rifampin. The detected synonymous mutation(s) do not confer resistance". "

(If other rpoB mutations have MDL interpretation "U" or "R", then report</td></tr>
</table>

[14] LIMS output was specifically designed for CalLIMS

			drug predicted susceptibility based on those mutations' interpretation)
		All other cases when S mutation detected in *rpoB* AND No other mutations interpreted as R or U are detected in *rpoB* (anywhere in the gene)	Type "Predicted susceptibility to rifampin"
WT (no mutations detected in the corresponding targets)			Type "No mutations associated with resistance to XXX detected"
"Insufficient Coverage" AND No other genes associated with given drug has mutation interpreted as R			Pending Retest (If other genes associated with the same drug have mutations with MDL interpretation "R", then report drug predicted susceptibility based on those targets)

5.2.1. The Drug interpretation is based only on the set of genes that is being reported. So for LIMS, if a gene is not on the reportable list, we should ignore it and only generate overall Drug interpretation based on mutations found (or not) in the genes listed in LIMS export.

5.2.1.1. As per section 4.2.2, if a mutation fails QC in the position (but breadth of coverage for locus is passing), it should be treated as "not real", i.e. the mutation should not be reported on LIMS report and will not be taken into consideration when determining the interpretation for the corresponding drug.

5.2.2. Column "M_DST_O01_Lineage" in LIMS report is populated from the main "Lineage" output of TB Profiler.

5.3. Outputting ***individual mutations*** in gene target fields in LIMS report (e.g. M_DST_B02_katG):

5.3.1. Only output mutations classified as "R" or "U" in "MDL interpretations" field of Laboratorian report, for the exception of RRDR region of *rpoB*:

5.3.1.1. Within *rpoB* codons 426-452, output ALL individual mutations (including synonymous);

5.3.1.2. After the synonymous mutation output text "[synonymous]", e.g.: "c.2630G>A (p.Asp877Asp) [synonymous]" (RIF interpretation output is changed accordingly in case if the only mutation detected in rpoB RRDR is synonymous; see above)

5.3.2. See instructions on output format for individual mutations below.

Information coming from Laboratorian report		**In LIMS output of TB Profiler**	
MDL Interpretation (for each given mutation)		**Gene Loci (e.g. M_DST_B02_katG), i.e. where individual mutations that are listed for each target**	
		Format	**Mutation should be listed in LIMS report**
R		Report the detected mutation in format "c.2066C>T (p.Ala689Val)" [15]	Yes
U		Report the detected mutation in format "c.2066C>T (p.Ala689Val)"	Yes
S	S for mutations in RRDR region of *rpoB* gene: If synonymous mutation present within *rpoB* codons 426-452 (P.S.: do not report synonymous mutations outside this region)	Report the detected synonymous mutation in format "c.2066C>T (p.Ala689Ala) [synonymous]" (Report whether synonymous mutation is detected alone in RRDR or there are other R mutations in RRDR)	Yes
	S- for all other cases	Mutation is not displayed on LIMS report. If the only mutation detected in given target gene was S, type "No high confidence mutations detected"	No
WT (no mutations detected in the corresponding target)		Type "No mutations detected" for the corresponding gene target	N/A
"Insufficient Coverage"		Type "No sequence" for the corresponding gene target (unless R mutation with adequate coverage detected in gene)	N/A

5.3.3. If a mutation in the same position (usually a deletion) is listed under different annotation types on Laboratorian report, use the one with highest read_support to decide which deletion to report in LIMS. Leave the alternative mutation annotation on Laboratorian report but do not use it for evaluation of predicted drug resistance.

[15] If multiple mutations in the gene, list separated by semicolon: "c.2066C>T (p.Ala689Val); c.169C>G (p.His57Asp)"

5.4. **MTBC ID by WGS only**: Parse TB Profiler original output that contains lineage prediction information to populate LIMS report field "M_DST_A01_ID" for species ID. Assign following values in M_DST_A01_ID depending on text in "main_lin" field of TBProfiler:

5.4.1.IF 90% of genes on the coverage report that are ALSO on the LIMS report have >=100% breadth of coverage at 10x, **THEN**

5.4.1.1. If main_lin field contains "lineage": "DNA of Mycobacterium tuberculosis species detected"

5.4.1.2. If main_lin field or sublin field contains "BCG": "DNA of Mycobacterium bovis BCG detected"

5.4.1.3. If those fields do **NOT** contain "BCG", but **DO** contain "bovis" or "La1": "DNA of Mycobacterium bovis (not BCG) detected"'

5.4.1.4. If main_lin field blank **OR** "NA" **OR** non-existent: "DNA of Mycobacterium tuberculosis complex detected"

5.4.2.ELSE: "DNA of Mycobacterium tuberculosis complex NOT detected"

5.4.3. It is acceptable to report resistance prediction result, as long as the following requirements are met:

5.4.3.1. Samples identified as MTBC by the pipeline (i.e. have any of the following WGS ID results: "DNA of Mycobacterium tuberculosis species detected", "DNA of Mycobacterium bovis BCG detected", "DNA of Mycobacterium bovis (not BCG) detected", or "DNA of Mycobacterium tuberculosis complex detected")

AND

5.4.3.2. Coverage of all the other gene markers reportable in WGS-DST assay is at least 90% (unless a deletion that would explain coverage <90% was detected automatically by the pipeline or upon SME review).

5.4.4. Samples that were identified as MTBC based on automated algorithm but have the rest of the gene targets covered at <90% breadth of coverage should be re-sequenced and, if coverage is not improved, reported as "DNA of Mycobacterium tuberculosis complex NOT detected".

5.5. **MTBC ID by tNGS only:** parse Coverage and Laboratorian reports to populate LIMS report field "M_DST_A01_ID" for species ID:

5.5.1.IF ≥ 90% of genes on the Coverage report that are ALSO on the LIMS report have ≥ 90% breadth of coverage at ≥10x, **AND:**

5.5.1.1. Laboratorian report for the given sample has mutation p.His57Asp (field tbprofiler_variant_substitution_aa) in *pncA* gene, **THEN** "M_DST_A01_ID" = "DNA of Mycobacterium bovis detected"

OR

5.5.1.2. Laboratorian report for the given sample does NOT have mutation p.His57Asp (field tbprofiler_variant_substitution_aa) in *pncA* gene, **THEN** "M_DST_A01_ID" = "DNA of Mycobacterium tuberculosis complex detected, not M. bovis"

5.5.2.IF < 90% of genes on the Coverage report that are ALSO on the LIMS report have ≥ 90% breadth of coverage at ≥10x, **THEN** "M_DST_A01_ID" = "DNA of Mycobacterium tuberculosis complex NOT detected"

6. Looker output

6.1.1. For Looker Drug interpretations (Looker matrix table), we should take into consideration all genes for the corresponding drugs, that are present in laboratorian report, because we will be importing information for all genes into the Looker.

6.1.2. "ID" field: copy over the value that is generated by TBP Parser for MTBC ID in the LIMS report (section 5.4 above).

6.1.3. "lineage" field: populate from main Lineage output of TB Profiler (section 5.2.2 above).

7. tNGS-specific analysis

7.1. QC regions for breadth of coverage calculations were determined by narrowing the regions covered by the primers to the regions that consistently obtained at least 20x depth – see section "Principles of Analysis" for the exact ranges.

7.1.1. An additional column ("Coverage_Breadth_R_expert-rule_region") is added to the coverage report that includes the breadth of coverage for the sites that either have "R" mutations or have an expert rule applicable. This column is not used for QC but is used for SME knowledge.

7.1.1.1. The "Percent_Coverage" column is renamed to "Coverage_Breadth_reportableQC_region"

7.1.1.2. The "Warning" column is renamed to "QC_Warning"

7.1.2. The primer regions for katG, and rrs were combined as they overlapped.

7.1.2.1. rpoB primer regions did not overlap and are present on the coverage report as "rpoB_1" and "rpoB_2"

7.1.2.1.1. Breadth of coverage QC for rpoB mutations uses **both** rpoB_1 and rpoB_2 and fails the mutation if at least one of the segments does not meet QC thresholds

7.2. Only the genes included in the tNGS assay are included in the LIMS report.

7.3. Certain sites in the tNGS assay are noisy and prone to false mutations. The ability to modify the minimum read support and frequency for those sites is available- see section "Principle of analysis" for details.

7.4. If a mutation appears outside the region covered by the primers:

7.4.1. Write "This mutation is outside the expected region" in the *Warning field*
AND
rationale = "NA"
confidence = "NA"
Looker interpretation = "NA"
MDL interpretation = "NA"
AND
Ignore this mutation in the LIMS report entirely