

# Detection of Mycobacterium tuberculosis mutations related to drug resistance combining ELITECH Group ELITE InGenius® Technology and MGI ATOplex DNBSEQ™ Technology

## ● Introduction

Tuberculosis (TB) is a preventable and transmittable disease. It is one of the top 10 causes of death worldwide and the world's second (after COVID-19) cause of death from a single infectious agent, ranking above HIV/AIDS. TB is caused by the bacillus Mycobacterium tuberculosis (MTB), which spreads through microdroplets expelled by affected people in the air, for instance when coughing. The disease primarily affects the lungs (pulmonary TB) but can also affect other organs or human tissues (extrapulmonary TB) in about 15% of cases. It is estimated that about a quarter of the world's population is infected with latent tuberculosis. The number of new TB cases has been decreasing slowly over the past years. In 2022, 10 million people developed tuberculosis, and out of these approximately 8% were people living with HIV.

When early detected and appropriately treated, TB is curable. However, globally in 2022, the total number of deaths caused by TB was 1.13 million in HIV-negative people and 167,000 in HIV-positive people. The World Health Organization (WHO) reported 19% net reduction in the global number of deaths caused by TB from 2015 to 2022. Though valuable, the result is still far from the milestone of a 75% reduction by 2025, set in the WHO's End TB strategy. The early diagnosis of TB, including universal drug-susceptibility testing and treatment of all people with TB are among the pillars of such WHO Strategy (WHO Global Tuberculosis Report 2022).

Drug-resistant tuberculosis (DR-TB) is a major challenge in the containment and eradication of TB. Drug resistance arises as a consequence of the use of anti-TB drugs, especially first-line drugs, like rifampicin and isoniazid. Rifampicin works by inhibiting bacterial DNA-dependent RNA polymerase, encoded by the *rpoB* gene (Hartmann 1967). Resistance to this drug has been associated mainly with mutations in a limited region of the *rpoB* gene (Telenti 1993). Multidrug-resistant TB (MDR-TB), which is resistant to both isoniazid and rifampicin, represents a serious concern in 2022, 410000 people were estimated to have developed MDR-TB or RR-TB.

Although culture is the gold standard test for diagnosis of TB, it requires more resources and takes longer to produce results (2–6 weeks) than molecular tests and sputum-smear microscopy, which can provide results in less than 1 day. Accurate and rapid detection of TB and MDR-TB is critical for improving patient care and decreasing disease transmission. Therefore, the WHO recommends using diagnostic molecular tests as the first approach for detecting TB among people with signs or symptoms. Nonetheless, despite the improvements, current molecular tests for MDR-TB can only detect a limited number of resistant mutations and their upgrade lies behind the continuous rise of new resistant mutations. In this picture, high throughput sequencing (NGS) is a technology that is gaining interest for the detecting multi and extensively drug-resistant (MDR-TB and XDR-TB) TB strains, since it provides the possibility to investigate a high number of sites within different genes carrying mutations that lead to resistance.

This pilot study presents preliminary data on a complete NGS workflow (Fig. 1) for TB resistances detection. It was designed to exploit the good synergy between two complementary solutions: ELITech Group MDR/MTB ELITE MGB® Kit assay on the ELITE InGenius integrated platform for nucleic acids preparation and the MGI ATOplex DNA Custom Library protocol analyzed with DNBSEQ™ sequencing technology:

- ELITech Group ELITE InGenius instrument is a sample-to-result technology able to perform nucleic acid extraction, Real-Time PCR (RT-PCR) using MDR/MTB ELITE MGB® Kit and results interpretation to identify MTB positive clinical samples (sputum) as a first diagnostic test.
- MGI ATOplex DNA Custom Library is a targeted multiplex PCR-based enrichment technique that, in combination with the high performance of DNBSEQ™ sequencing technology, allows interrogate simultaneously whole MTB genes sequences to better characterize the genotypic drug-resistance also for mutation unknown.

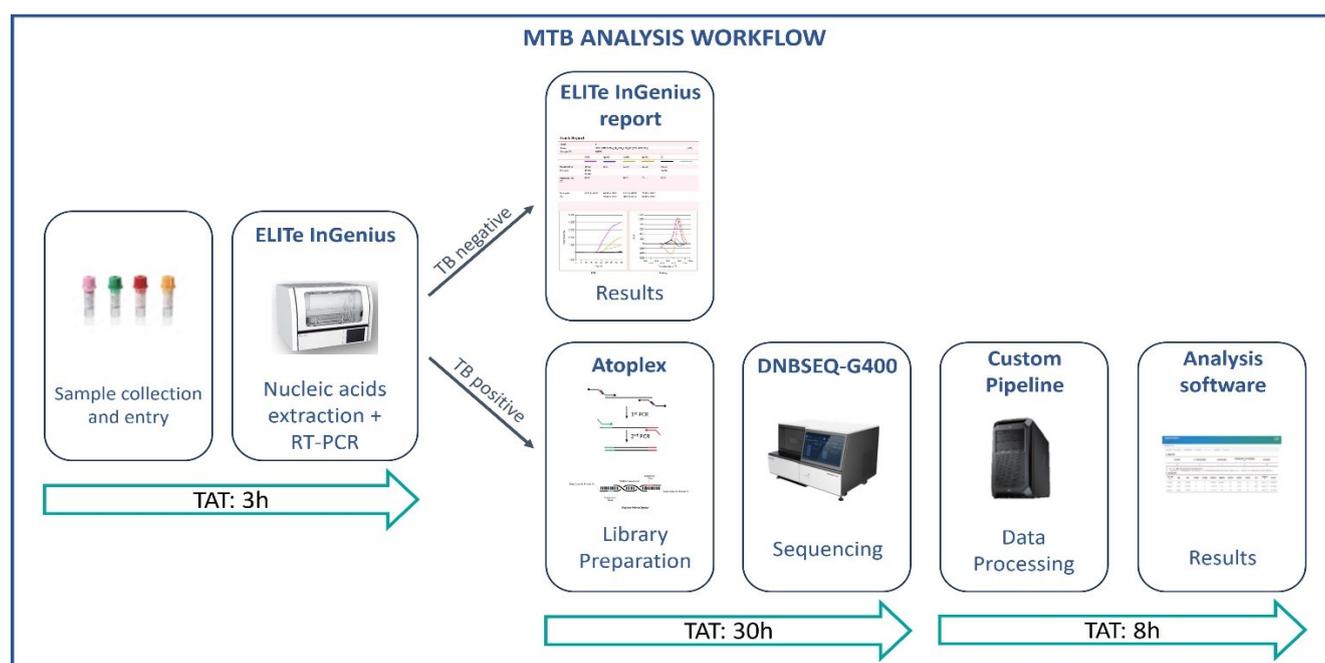
This all-in-one workflow could be easily adapted and validated at any laboratory, to obtain a rapid and multi comprehensive solution for the detection of TB drug resistance.

## Experimental Methods



### ELITE InGenius – ATOplex DNBSEQ™ Workflow

The complete workflow allows an initial identification of MTB-positive sputum using MDR/MTB ELITE MGB® Kit Real-Time PCR assay on the automated sample-to-results instrument ELITE InGenius. Positive samples are subsequently analyzed with a downstream process of targeted analysis by amplicon sequencing using the MGI ATOplex DNA Custom Library technology. The samples are then sequenced on the MGI DNBSEQ™ sequencing platform. The data are finally analyzed through a bioinformatics pipeline to detect potential mutations in different target genes that might be implicated in resistance to pharmacological treatments against M. tuberculosis (Fig.1).



**Figure 1:** Complete workflow created from nucleic acid extraction, Real-Time PCR analysis on the ELITE InGenius®, NGS with the DNBSEQ™ technology, and data analysis with custom bioinformatic pipeline.



## Samples Preparation

Different sample types were used to test the whole process. Commercial M. Tuberculosis H37Rv genomic DNA (ATCC) was used as a positive control. Ultrapure water was used as a negative control. Simulated samples were prepared by adding inactivated MTB isolates obtained from external providers at concentrations of 10<sup>5</sup> CFU/mL and 10<sup>4</sup> CFU/mL to MTB-negative sputum specimens. Two different strains were tested: one wild type (H37Rv) and one carrying a mutation on the katG gene (codon 315, responsible for isoniazid resistance). Moreover, three inactivated clinical isolates from TB-positive sputum samples obtained from external providers were also processed. The three samples were characterized using MDR/MTB ELITE MGB<sup>®</sup> Kit on ELITE InGenius<sup>®</sup>, one was detected as MTB wild-type, while the other two were detected as MTB mutated respectively for katG and rpoB (codon 526, responsible for rifampicin resistance). In total, 11 samples plus a positive control were tested to evaluate the whole process for sequencing analysis.

Sample No.:	Samples	Concentration	Specimen
1	Positive control 50 ng DNA MTB	n.a.	n.a.
2	Simulated sample – MTB WT replicate 1	10 <sup>5</sup> CFU/mL	SPUTUM
3	Simulated sample – MTB WT replicate 2	10 <sup>5</sup> CFU/mL	SPUTUM
4	Simulated sample – MTB WT replicate 1	10 <sup>4</sup> CFU/mL	SPUTUM
5	Simulated sample – MTB WT replicate 2	10 <sup>4</sup> CFU/mL	SPUTUM
6	Simulated sample – MTB MUT katG replicate 1	10 <sup>5</sup> CFU/mL	SPUTUM
7	Simulated sample – MTB MUT katG replicate 2	10 <sup>5</sup> CFU/mL	SPUTUM
8	Simulated sample – MTB MUT katG replicate 1	10 <sup>4</sup> CFU/mL	SPUTUM
9	Simulated sample – MTB MUT katG replicate 2	10 <sup>4</sup> CFU/mL	SPUTUM
10	Isolated sample – MTB MUT katG	n.a.	SPUTUM
11	Isolated sample – MTB MUT rpoB	n.a.	SPUTUM
12	Isolated sample – MTB WT	n.a.	SPUTUM

**Table 1:** Samples List with number of replicates, concentrations, and type of specimen.

## DNA extraction and Real-Time PCR analysis

Samples DNA extraction and Real-Time PCR analysis to detect MTB positivity were performed using the ELITE InGenius® platform and MDR/MTB ELITE MGB® Kit. Through a preload assay, the instrument extracts DNA from fluidified, decontaminated, and inactivated sputum samples and subsequently performs the Real-Time PCR analysis to detect the presence of MTB (IS6110 target gene). The process uses just a minimal part of the extracted DNA sample (40 µl out of 100 µl of total eluate volume) and after the extraction, the remaining eluate volume was utilized to prepare sequencing libraries for the DNBSEQ™ sequencing.

## Library preparation for sequencing

Sequencing libraries were prepared using the MGI ATOplex DNA Custom Library kit, which contains all reagents and controls required for the optimized ATOplex DNBSEQ™ workflow. The process starts with a first PCR step to amplify the target gene regions, and a second PCR step to amplify the dual-barcode bacterial genome. In this study, 5 target genes associated with resistance to first-line drugs were investigated. The MTB strain present in the sample was predicted to be susceptible or resistant to each antibiotic or with yet-to-be-characterized mutations, based on the absence or presence of mutations in these loci, the target gene region and interrogation of reference databases. For each gene, the full protein-coding regions and some upstream and downstream bases were covered to maximize the discovery of known and novel variants (Table 2).

All target regions were covered, and 116 primer pairs were designed.

Gene Target	Function	Major Drug Resistance	Gene size covered (bp)	Size of Full Genes Covered (bp)	Start position	End position
rpoB	DNA-directed RNA polymerase β-chain	Rifampicin	3518	3552	759807	763325
katG	Catalase-peroxidase-peroxynitritase T	Isoniazid	2222	2413	2153889	2156111
inhA	NADH-dependent enoyl-[acyl-carrier-protein] reductase	Isoniazid	809	1010	1674202	1675011
gyrA	DNA gyrase subunit A	Fluoroquinolones	2516	2552	7302	9818
embB	Membrane indolylacetylinsitol arabinosyltransferase B	Ethambutol	3296	3321	4246514	4249810

**Table 2:** Summary of target genes, function, major drug resistance and gene size covered (bp) with the MGI ATOplex DNA Custom library protocol.

After the dsDNA libraries preparation, the specific steps to MGI's DNBSEQ™ technology were followed. At first, the quality of the dsDNA libraries was evaluated using the Agilent TapeStation platform. All samples showed good quality and were used to perform the circularization of single-strand DNA to produce the ssCircDNA libraries. The ssCircDNA libraries obtained were then converted to DNA Nanoballs (DNB) for sequencing.

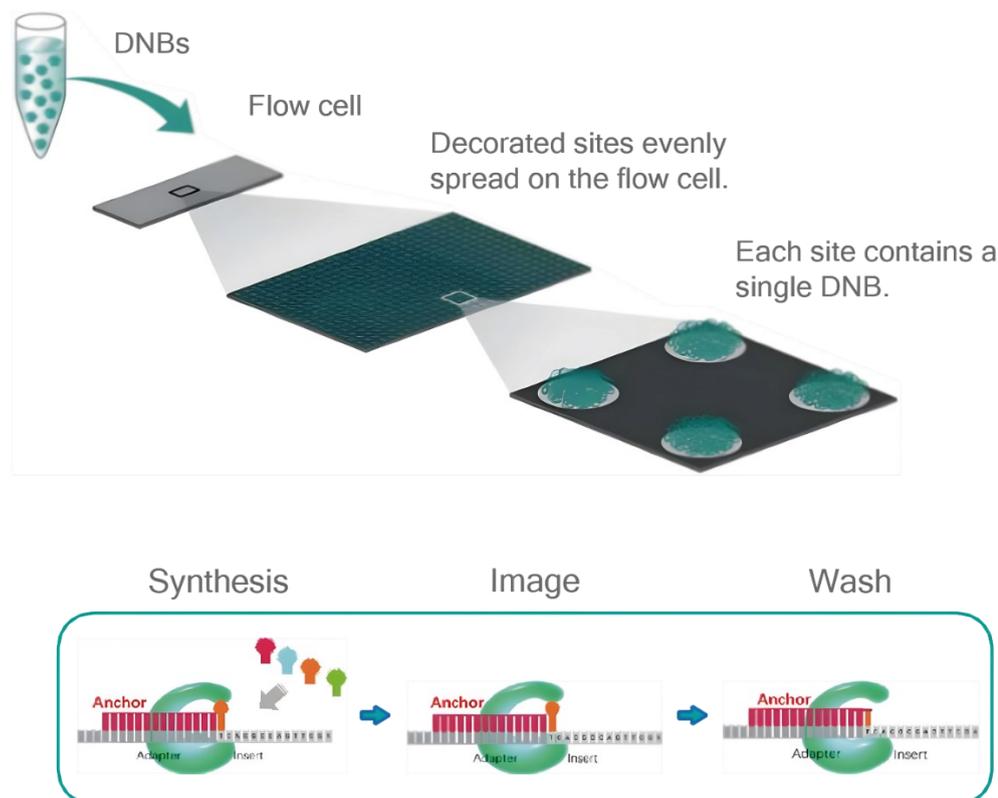
A negative control (ultrapure water) sample was introduced during the MGI ATOplex DNA Custom Library preparation as a cross-contamination control during this step.

## Sequencing

DNB sequencing takes place on high-density pattern arrays that immobilize individual DNBs for highly parallelized tracking of dNTP incorporation during strand extension (Figure 2). Sequencing starts with the hybridization of DNBs to anchor spots. After primer hybridization, the flow cell is flushed with fluorescently labeled dNTP probes. Unbound probes are washed away, and bound probes are stimulated to fluoresce. High resolution imaging and proprietary algorithms transform signals into high-quality and highly accurate sequencing results.

Libraries were pooled and sequenced using the DNBSEQ-G400 sequencer and the FCL PE100 sequencing set

with dual barcode sequencing, according to the instructions of the ATOplex protocol. Four pools of samples were sequenced in each run. The four pools were loaded manually into the four lanes of the flow cell using the MGIDL-200H Portable DNB Loader since they shared the same barcodes.



**Figure 2:** DNB sequencing uses the combinatorial probe anchor system (cPAS) on a patterned array flow cell to provide high sequencing accuracy with improved imaging and reduced index hopping.

## • Data analysis and sequencing assembly

The sequencing data quality control check was performed using the FastQC software. SOAPnuke software was used to remove low-quality reads and adapter sequences from the Fasta output file. Minimap2 software was used to align reads to the MTB H37Rv reference genome (NC\_000962.3), and then variants were called using the GATK HaplotypeCaller software. BAM file quality and mapping statistics were assessed using Bamstats software. Data generated was visualized using IVG software.

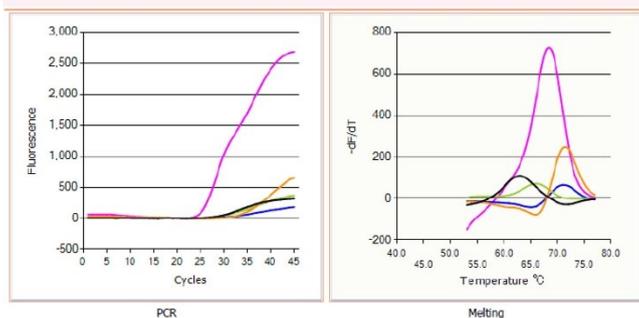
## • Results

All samples prepared were successfully processed with ELITech Group ELITe InGenius® platform and MDR/MTB ELITe MGB® Kit, as confirmed by the amplification of the internal Control of the assay. Analysis of the real-time PCR amplification and melting curve analysis confirmed the expected MTB genotype. An example of Real-Time PCR output from ELITe InGenius® is shown in Fig.3.

The remaining MTB DNA eluates from all samples were used to prepare sequencing libraries on the DNBSEQ-G400 sequencer. The libraries displayed good quality and concentration, resulting in appropriateness for the sequencing analysis. All samples were analyzed together in a single pool in 4 replicates across 4 different sequencing lanes of the DNBSEQ-G400 flowcell.

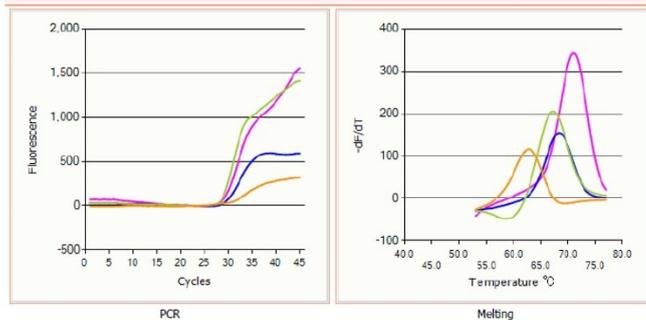
## Track Report

Track	5				
Assay	MDR-MTB ELITE_SP_200_100_05 [TB1 PCR Mix] (IVD)				
Sample ID	MTB				
	MTB	rpoB3	rpoB4	rpoB2	IC
Measured Ct	24.33	32.20	29.15	32.22	29.93
Ct Limits	45.00 31.00				34.00
Measured Tm (°C)	68.5	71.2	66.0	71.6	62.9
Tm Limits (°C)	67.0 to 69.9	68.0 to 80.0 50.0 to 67.9	63.5 to 80.0 50.0 to 63.4	70.0 to 80.0	



## Track Report

Track	6			
Assay	MDR-MTB ELITE_SP_200_100_05 [TB2 PCR Mix] (IVD)			
Sample ID	MTB			
	katG	inhA	rpoB1	IC
Measured Ct	28.50	29.06	27.71	31.16
Ct Limits				34.00
Measured Tm (°C)	71.0	68.5	67.3	62.7
Tm Limits (°C)	70.0 to 80.0 50.0 to 69.9	66.0 to 80.0 50.0 to 65.9	66.0 to 80.0 50.0 to 65.9	



**Figure 3:** Example of Real-Time PCR results obtained with the ELITE InGenius® instrument.

The sequencing process generated an output of 386M clean reads for all the samples analyzed, achieving a Phred Quality Score of 30 (Q30) of 95% or higher. This demonstrates how the complete process was able to produce a considerable amount of data with an excellent quality of base call accuracy, as most of the reads obtained were able to reach a Q30 score. In addition, 128M reads out of the 386M reads obtained were successfully mapped to the expected rpoB, katG, inhA, gyrA and embB gene targets (Table 3).

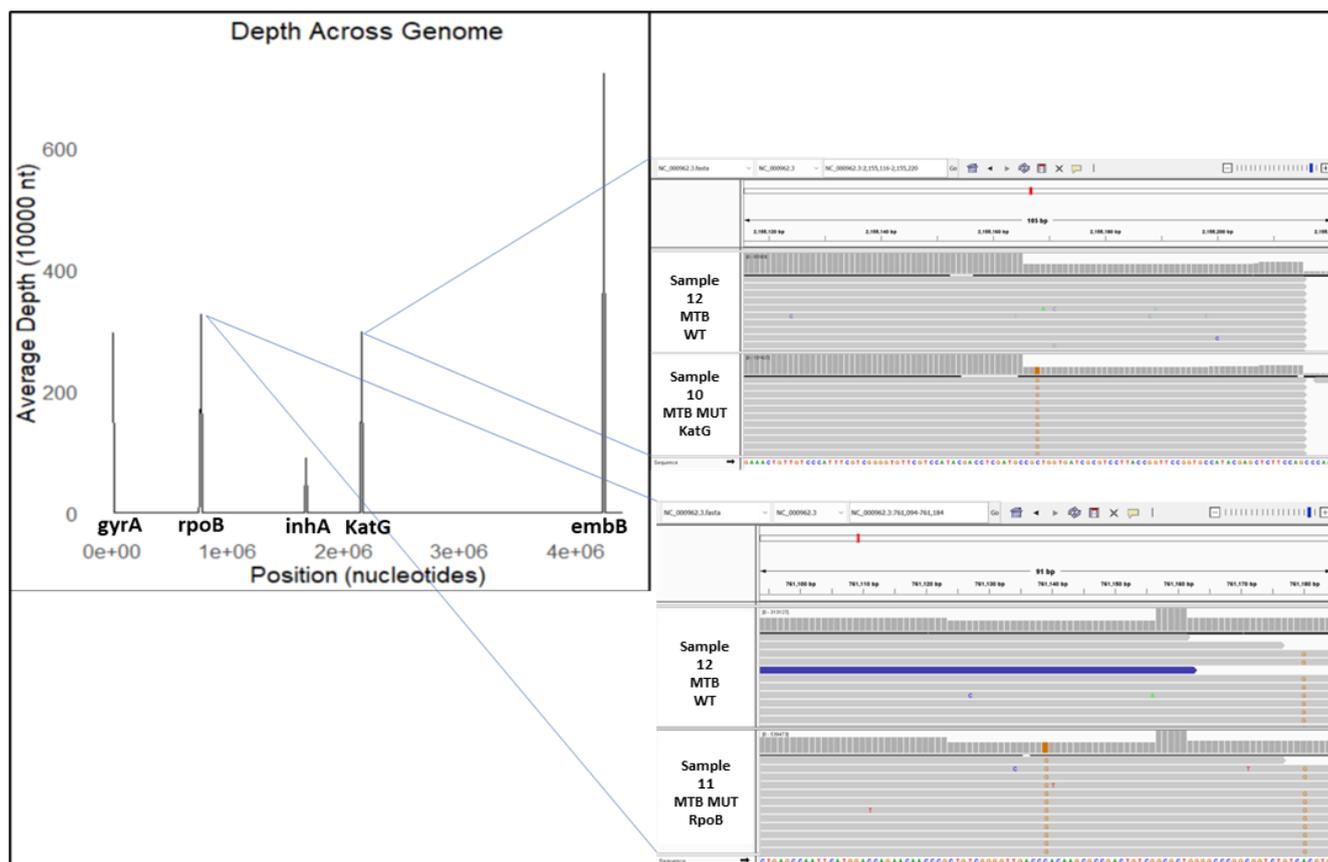
Sample No.:	Type of Sample	Specimen	Total number of clean reads	Mapping reads	Q30% sample
1	Positive Control 50 ng DNA MTB	/	33955256	33886528	94,09
2	MTB_WT 10 <sup>5</sup> CFU/mL - Rep1	SPUTUM	60805390	12333839	96,81
3	MTB_WT 10 <sup>5</sup> CFU/mL - Rep2	SPUTUM	27910758	3662780	97,00
4	MTB_WT 10 <sup>4</sup> CFU/mL - Rep1	SPUTUM	31910826	174934	97,29
5	MTB_WT 10 <sup>4</sup> CFU/mL - Rep2	SPUTUM	33366572	184129	96,68
6	MTB_katG 10 <sup>5</sup> CFU/mL - Rep1	SPUTUM	31001452	6772865	96,86
7	MTB_katG 10 <sup>5</sup> CFU/mL - Rep2	SPUTUM	26798066	7751015	96,41
8	MTB_katG 10 <sup>4</sup> CFU/mL - Rep1	SPUTUM	33163982	1036209	96,86
9	MTB_katG 10 <sup>4</sup> CFU/mL - Rep2	SPUTUM	29130560	1280122	96,18

10	Isolated_sample - MTB MUT katG	SPUTUM	24669312	24613955	95,36
11	Isolated_sample – MTB MUT rpoB	SPUTUM	24715790	24619025	95,55
12	Isolated_sample – MTB WT	SPUTUM	29269410	12405084	96,25
<b>Total</b>	<b>386697374</b>	<b>128720485</b>			

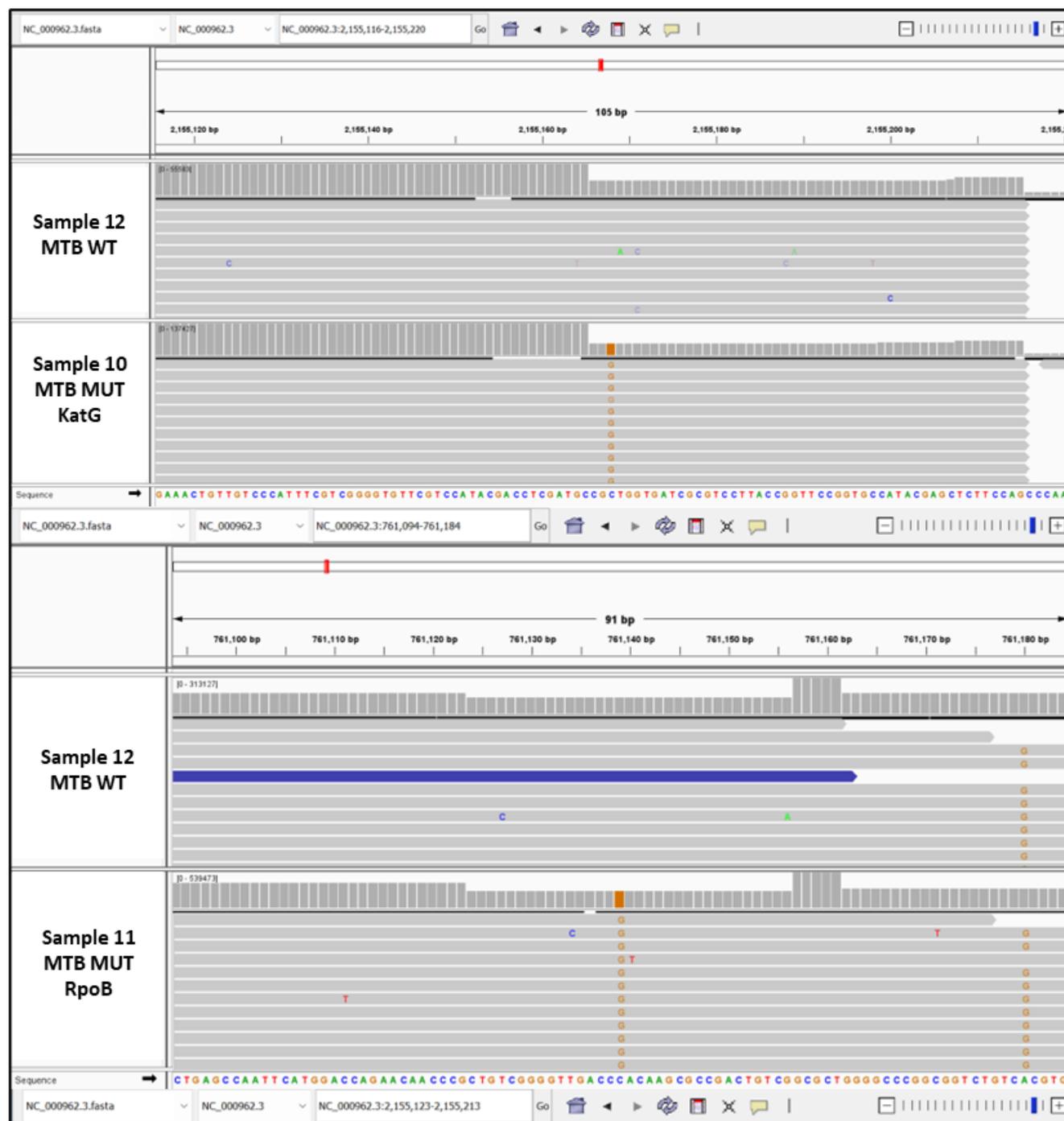
**Table 3:** Overall total reads count and Q30% for the samples analyzed.

An overview of the overall average depth obtained for all samples analyzed across the MTB genome is represented in Figure 3. The data obtained from the targeted sequencing process fell in the genome region of the 5 target genes as expected, moreover the number of reads generated was sufficient to detect the presence of mutations with extremely high precision. Mutations in *katG* and *rpoB* genes were properly detected and visualized with the IVG genome browser as shown in Fig.4 A and B.

A)



B)



**Figure 4:** Example of sequencing data distribution across the MTB genome and of a visualization of mutation on *katG* and *rpoB* target genes. A) Each peak shows the sequencing depth of each target gene of the panel tested across the MTB genome. B) Extract of the alignment of the sequencing data on the IVG genome browser and visualization of the mutations detected.

The analysis with GATK HaplotypeCaller allowed the precise detection of the genotype of each one of the 11 samples analyzed with the MGI ATOplex protocol. All simulated MTB wild-type samples (samples 2-5) were confirmed at different concentrations and replicates, likewise, simulated MTB-resistant samples (samples 6-9) were properly detected and the expected mutation pattern was confirmed across different concentrations and replicates. Similarly, the sequencing analysis allowed the identification of the expected genotype in the three clinical isolates (samples 10-12) and the identification of new unknown mutations not reported as related to

drug resistance. A summary of the results obtained is reported in Table.4.

Sample No.:	Sample name	Matrix	Ref Genome	POS	Gene	REF	ALT	QUAL
1	Positive control 50 ng DNA MTB	/	NC_000962.3	/	/	/	/	/
2	Simulated sample MTB_WT 10e5 CFU/mL - Rep1	SPUTUM	NC_000962.3	/	/	/	/	/
3	Simulated sample MTB_WT 10e5 CFU/mL - Rep2	SPUTUM	NC_000962.3	/	/	/	/	/
4	Simulated sample MTB_WT 10e4 CFU/mL - Rep1	SPUTUM	NC_000962.3	/	/	/	/	/
5	Simulated sample MTB_WT 10e4 CFU/mL - Rep2	SPUTUM	NC_000962.3	/	/	/	/	/
6	Simulated sample MTB_KatG 10e5 CFU/mL - Rep1	SPUTUM	NC_000962.3	7362	gyrA	G	C	14662,04
				7585	gyrA	G	C	6889,04
				9304	gyrA	G	A	8838,04
				9596	gyrA	G	T	3584,04
				762434	rpoB	T	G	9983,04
				763031	rpoB	T	C	12109,04
				2154724	katG	C	A	2638,04
				<b>2155168</b>	katG	C	G	2230,04
7	Simulated sample MTB_KatG 10e5 CFU/mL - Rep2	SPUTUM	NC_000962.3	7362	gyrA	G	C	12677,04
				7585	gyrA	G	C	7974,04
				9304	gyrA	G	A	9481,04
				9596	gyrA	G	T	2505,04
				762434	rpoB	T	G	10778,04
				763031	rpoB	T	C	12952,04

				2154724	katG	C	A	2563,04
				<b>2155168</b>	katG	C	G	2644,04
<b>8</b>	Simulated sample MTB_KatG 10e4 CFU/mL - Rep1	SPUTUM	NC_000962.3	7362	gyrA	G	C	6113,04
				7585	gyrA	G	C	6572,04
				9304	gyrA	G	A	5542,04
				9596	gyrA	G	T	48,04
				762434	rpoB	T	G	4796,04
				763031	rpoB	T	C	6708,04
				2154724	katG	C	A	2096,04
				<b>2155168</b>	katG	C	G	2028,04
<b>9</b>	Simulated sample MTB_KatG 10e4 CFU/mL - Rep2	SPUTUM	NC_000962.3	7362	gyrA	G	C	6701,04
				7585	gyrA	G	C	4330,04
				9304	gyrA	G	A	5566,04
				9596	gyrA	G	T	89,04
				762434	rpoB	T	G	5317,04
				763031	rpoB	T	C	6545,04
				2154724	katG	C	A	2070,04
				<b>2155168</b>	katG	C	G	1968,04
<b>10</b>	Isolated sample MTB_KatG	SPUTUM	NC_000962.3	7362	gyrA	G	C	19866,04
				7585	gyrA	G	C	15506,04
				9304	gyrA	G	A	18701,04
				763031	rpoB	T	C	22912,04
				2154724	katG	C	A	5456,04

				<b>2155168</b>	katG	C	G	4195,04
<b>11</b>	Isolated sample MTB_RpoB	SPUTUM	NC_000962.3	7362	gyrA	G	C	22349,04
				7585	gyrA	G	C	17180,04
				9304	gyrA	G	A	17028,04
				<b>761139</b>	rpoB	C	G	15605,04
<b>12</b>	Isolated sample MTB_WT	SPUTUM	NC_000962.3	7362	gyrA	G	C	16243,04
				7585	gyrA	G	C	12273,04
				9214	gyrA	C	T	3879,04
				9304	gyrA	G	A	13486,04
				760115	rpoB	C	T	3324,04

**Table 4:** Summary of the detected variant call. For each sample, the detected mutations are listed according to position on the MTB genome, gene name, reference/alternate base and quality score called by GATK HaplotypeCaller. Expected mutations of each sample are reported in bold. Additional unknown mutations were also detected during the analysis.

## Conclusion

This preliminary study shows the applicability of a time-to-result-effective workflow for NGS sequencing, combining Elitechgroup ELITech Group ELITE InGenius® system for extraction and Real-Time PCR analysis and together with MGI ATOplex DNBSEQ™ sequencing technology. This workflow potentially constitutes a solution to investigate the genotypic drug resistance of M. tuberculosis, providing a substantial improvement over the current MTB diagnosis. Compared to other systems available on the market, the combination of the ELITE InGenius®' sample-to-results technology with ATOplex's targeted sequencing technology, substantially reduces the time and cost of analysis of clinical MTB samples, limiting at the same time the manual activity of the operator.

This newly established process may allow the identification of multiple resistant mutations on samples at different bacillary titres, on the same sample used for the diagnosis of TB, without the need for a second specimen. The complete workflow can be completed in 48 hours and it allows the identification of samples positive for MTB via Real-time PCR analysis in 3 hours using the MDR/MTB ELITE MGB® Kit. The leftover eluate from the same samples collected after Real-Time PCR analysis can be used to perform downstream targeted sequencing for genotypic mutation identification, thus keeping the amount of biological sample to be taken from the patients limited.

With the experiment performed the leftover of the MTB DNA extracted with the ELITE InGenius® was subsequently used to perform a targeted sequencing analysis using the MGI ATOplex DNA Custom library preparation protocol and sequenced with DNBSEQ-G400 sequencing instrument. The whole process yielded a consistent amount of high-quality data. Cleaned data were used to analyze the genotype of the genes targeted

with MGI ATOplex protocol, as a final result the expected genotype has been confirmed for all samples. The sequencing process was able to provide details of the single base variants detected in the MTB mutated samples.

The ATOplex sequencing DNA Custom Library protocol was developed and designed to cover the entire sequence of the genes that are implicated in the drug-resistance mechanism of MTB, allowing in a single assay the analysis of a large number of codons known to be implicated in drug treatment resistance when carrying mutations. In addition, this new method also allows the detection of potential new mutations on the target genes that can be monitored to evaluate their possible link to unexpected events of drug resistance not associated with known mutations.

Further investigations are required to assess the whole process performances, such as the limit of sensibility, the precision of analysis with additional mutations (*inhA*, *gyrA*, *embB*) and the performances of the whole process in combination with clinical samples. Also, the applicability to other sequencing systems, like the smaller platforms DNBSEQ-G99 and DNBSEQ-E25, should be tested. The performance evaluation will be crucial for the usability of the workflow for diagnostic purposes. Nonetheless, the easiness of use and the reduced turnaround time to result constitute an unrivalled advantage in disseminating the use of NGS into molecular diagnostics laboratories.

## References

1. World Health Organization (WHO) Global tuberculosis report 2022.
2. Hartmann G, Honikel KO, Knüsel F, Nüesch J. The specific inhibition of the DNA-directed RNA synthesis by rifamycin. *Biochim Biophys Acta*. 1967;145(3):843-4. doi: 10.1016/0005-2787(67)90147-5. PMID: 4863911.
3. Telenti A, Marchesi F, Balz M, Bally F, Böttger EC, Bodmer T. Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. *J Clin Microbiol*. 1993 Feb;31(2):175-8. doi: 10.1128/jcm.31.2.175-178.1993. PMID: 8381805; PMCID: PMC262730.

## Ordering information

Product Name	Catalog Number
ELITe InGenius®	INT030 (Contact Elitech)
Genetic Sequencer DNBSEQ-G400	900-000170-00
ELITe InGenius® SP 200 Extraction Kit	INT032SP200 (Contact Elitech)
MDR/MTB ELITe MGB® Kit	RTS120ING (Contact Elitech)
MDR/MTB-ELITe Positive Control	CTR120ING (Contact Elitech)
ATOplex Custom Panel	1000025243

ATOPlex DNA Universal Library Prep Module	1000021191
ATOPlex Single Barcode Primer Module (01-96)	1000024934
DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE100)*	1000016950
MGIEasy Dual Barcode Circularization Kit	1000020570
CPAS Barcode Primer 3 Reagent Kit DA	1000020834

*\*Unless otherwise informed, this StandardMPS sequencing reagent is not available in Germany, UK, Sweden, and Switzerland*