SUPPLEMENTARY MATERIAL

Accuracy of an Amplicon Sequencing Nanopore Approach to Identify Variants in Tuberculosis Drug Resistance-associated Genes

Supplementary Methods

1) Optimization of the single-tube multiplex PCR reaction

A qPCR of the PCR product was performed to set up the optimum concentration of each primer to obtain an equimolar amount of all the amplicons. The qPCR reaction volume was 20uL containing 1X volume Kapa SYBR Fast qPCR master mix (KAPA Biosystems, Roche), 0.2uM primers and 0.2ng of DNA (PCR amplicons). Water was added to reach the final volume. For this test, reverse primers were designed to amplify a small region of each gene (150-200bp) checking the specificity to avoid non-specific interactions (See **Table S6** containing the list of the qPCR primer sequences).

2) Barcodes design

We designed a set of 20 different barcodes, maximizing the genetic distance between them and supporting the correction of insertions, deletions and substitutions. We used the R package DNABarcodes to create barcodes of 36 bp using the Sequence-Levenshtein metric under the Ashlock heuristic. Each barcode consisted of a unique region of 20bp and two common regions of 8 bp (ACGATCTA) flanking this central sequence (See **Table S2** containing customized barcode sequences)**.** Primers with the barcode sequences and their complementary strands were ordered from IDT with HPLC purification.

Barcodes were designed using a R package called DNABarcodes using the Sequence-Levenshtein metric under the Ashlock heuristic model (1):

```
mySeqlevSet <- create.dnabarcodes(10, metric="seqlev", euristic="ashlock")
```

This created 10 bp barcodes. We built our 20 bp barcodes by joining two different barcodes of 10 bp.

3) Enrichment PCR step for lower yield samples

An additional PCR step was performed after the ligation of barcodes, only in samples that didn't yield less than 500ng of gene product in the first PCR. The reaction mix was prepared in a final volume of 50uL containing 10uL HiFi GC buffer 5X, dNTPs 0.3mM each, 0.4U of Kapa HiFi HS polymerase (Kapa Biosystems®), 5uL of each forward barcode 10uM were added to its corresponding sample, 2ng of barcoded DNA and water until reaching the final volume. Thermocycling conditions consist on a denaturation step at 95ºC during 3 minutes followed by 15 cycles of amplification as follows: 20 seconds at 98ºC, 15 seconds at 65ºC (primers annealing), 2 minutes at 72ºC (primers extension) and 5 minutes at 72ºC (final extension step). Final PCR enriched product was purified with 0.6X volume AMPure XP magnetic beads (following the Agencourt AMPure XP purification protocol) and eluted from the beads with Tris 10mM (pH=8.5) and quantified with Qubit®.

4) Modifications in the MinION library preparation protocol

Incubation time was increased from 10 to 30 minutes in the dA-tailing step, and also in the adapter ligation step to 30 minutes.

5) Base calling

We run guppy basecaller using 'flip flop' algorithm (2) by applying the following parameters:

```
guppy_basecaller --input_path $INPUT \
      --save_path $OUTPUT\
      --verbose_logs \
      --cpu_threads_per_caller 1 \
      --num_callers 20 \
      --config dna_r9.4.1_450bps_flipflop.cfg
```

6) Demultiplexing

To demultiplex the multi-fastq obtained after performing the base calling we run porechop (https://github.com/rrwick/Porechop):

```
python porechop-runner.py -i $FASTQ -b demultiplex/
```

This python script takes a multifasta file containing all the reads and classifies the reads depending on their barcode sequence generating a folder with separated fasta files, one per

sample. We modified the Oxford Nanopore barcode sequences by our customized barcode sequences. To demultiplex the reads, default parameters were used: --check-reads 10000, --adapter_threshold 90, --scoring_scheme (match = 3, mismatch = -6, gap open = -5, gap extend = -2), --end_size 150, --min_trim_size 4, --ehd_threshold 75, --extra_end_trim 2, --middle_threshold 85, --min_split_read_size 1000, --extra_middle_trim_good_side 10, --extra_middle_trim_bad_side 100.

7) Mapping

Mapping was done aligning reads to a multi fasta file containing the sequences of the genome regions included in the panel (extracted from the reference strain NC_000962.3 (3)) using minimap2 (4) with the following parameters:

```
# Mapping reads to the reference
minimap2 -ax map-ont $REFERENCE $FASTQ -t20 > $SAM

# Filtering mapped reads by mapping quality
awk '$1 ~ /^@/ || $5 == 60' $SAM > $MAPQ60.sam

# Sorting sam files
samtools sort $MAPQ60.sam > $MAPQ60.sort.sam

# Indexing sam files
samtools view -b ${sample}.MAPQ60.sort.sam -T $REFERENCE > $MAPQ60.sort.bam
```

8) MinION Variant Calling

Obtention of mpileup files (5):

```
samtools mpileup -AB -d 1000000 -f $REFERENCE $sort.bam > $MPILEUP
```

To call SNPs, we execute VarScan (6) on the mpileup files with the following parameters for samples sequenced with MinION: minimum read depth = 50, minimum number of reads supporting a position = 2, minimum base quality at a position to count a read = 10, minimum variant allele frequency threshold = 0.40 and minimum frequency to call homozygote = 0.90.

```
java -Xms10G -Xmx32G -jar VarScan.v2.3.7.jar pileup2snp $MPILEUP --min-coverage 50 --min-reads2 2 --min-avg-qual 10 --min-freq-for-hom 0.90 --min-var-freq 0.40 > $SNP
```

To call Insertions/Deletions (indels) we execute VarScan (6) on the mpileup files with the following parameters for samples sequenced with MinION:

```
java -Xms10G -Xmx32G -jar VarScan.v2.3.7.jar pileup2indel $MPILEUP --min-coverage 50 --min-reads2 2 --min-avg-qual 15 --min-freq-for-hom 0.90 --min-var-freq 0.1 --p-value 99e-02 > $INDEL
```

9) Calibration of MinION Variant Calling

Formulas to evaluate True positive rate (TPR), true negative rate (TNR) or recall, precision, agreement, F1-score, accuracy and error rate (7,8):

$$True\ Positive\ Rate\ = \frac{TP}{TP\ +\ FN}$$

$$True\ Negative\ Rate\ = \frac{TN}{TN\ +\ FP}$$

$$Precision\ = \frac{TP}{TP\ +\ FP}$$

$$\%\ Agreement\ = \frac{TP}{TP\ +\ FP\ +\ FN}\cdot 100$$

$$F1-Score\ = \frac{2\ \cdot\ (TPR\ \cdot\ Precision)}{TPR\ +\ Precision}$$

$$Accuracy\ (\%)\ = \frac{TP\ +\ TN}{TP\ +\ TN\ +\ FP\ +\ FN}\cdot 100$$

$$Error\ Rate\ MinION = \frac{FP\ +\ FN}{Total\ SNPs\ MinION}$$

SUPPLEMENTARY TABLES

Table S1: Primer sequences and their corresponding final concentrations for PCR

Primer	Sequence	PCR Concentration (uM)
katG_F_1	TGCGGCGGGTTGTGGTTGAT	0.1
katG_R_2	GCGCTACGAGTCCAGGGTCG	0.1
gyrA_F_1	AGATGACAGACACGACGTTGCC	0.2
gyrA_R_1	AGCCTAGCTGCCCGATTCCT	0.2
inhA_F_2.1	TTGGCGCCATGGAAGGCAGA	0.1
inhA_R_2.1	TGGCTAGTCGAGCGAACCGC	0.1
gyrB_F_4	TGCGGTTGGCGGCCTATCAA	0.2
gyrB_R_1	CGCAGGGTTGCGTTAGACATCC	0.2
pncA_F_1	ATCCGGACACTTGCCACCCG	0.2
pncA_R_1	GTTGGCGTTGCGTTGGGTCC	0.2
rpoB_F_int	GTCGCATGAAGTGCTGGAAGG	0.3
rpoB_R_int	GAAGTTGACGTCGAGCACGTAAC	0.3
embB_F_1	CCAGACGCCGTTGTCGAGGA	0.6
embB_R_2	GGCCTGGTGCATACCGAGCA	0.6
eis_F_1	AGGGTGACGAGTCCTGGGGT	0.2
eis_R_1	GAAGCGATGAGGTGGGGGCA	0.2
rrs_F_1	ACGAGCGTCCGAAGGCTGTC	0.2
rrs_R_2	GCCATCACCACCCTCCTCCG	0.2
L125_F_3	ACCCGCACTATGCCTGGCTG	0.4
L125_R_3	TGGATGGCGCTCAACGGGAG	0.4
L346_F_1	TCCCGACGGTGCGTGACTTG	0.4
L346_R_3	CGGCAGTGCCAGTTCATGCC	0.4

Table S2: Barcode sequences

Barcode		Sequence
BC01	Barcode_2_263_Fw	ACGATCTACCGCACTGAATCCAACACAAACGATCTAT
	Barcode_2_263_Rv	TAGATCGTTTGTGTTGGATTCAGTGCGGTAGATCGT
BC02	Barcode_3_262_Fw	ACGATCTAAGTCTACGGTGAGTTGACAAACGATCTAT
	Barcode_3_263_Rv	TAGATCGTTTGTCAACTCACCGTAGACTTAGATCGT
BC03	Barcode_1070_914_Fw	ACGATCTAGTGACCGCTACAGCTCATCAACGATCTAT
	Barcode_1070_914_Rv	TAGATCGTTGATGAGCTGTAGCGGTCACTAGATCGT
BC04	Barcode_1313_1025_Fw	ACGATCTAATGTGGTCGGTGATACCGTAACGATCTAT
	Barcode_1313_1025_Rv	TAGATCGTTACGGTATCACCGACCACATTAGATCGT
BC05	Barcode_201_70_Fw	ACGATCTATAGTGGTGAACGGTCTAGAAACGATCTAT
	Barcode_201_70_Rv	TAGATCGTTTCTAGACCGTTCACCACTATAGATCGT
BC06	Barcode_1400_2062_Fw	ACGATCTATAGGATCTGGCAGGTTGTCTACGATCTAT
	Barcode_1400_2062_Rv	TAGATCGTAGACAACCTGCCAGATCCTATAGATCGT
BC07	Barcode_565_556_Fw	ACGATCTAGACTTGCAGAGCTTATGAGAACGATCTAT
	Barcode_565_556_Rv	TAGATCGTTCTCATAAGCTCTGCAAGTCTAGATCGT
BC08	Barcode_579_1859_Fw	ACGATCTATCAGACTAGAACAACGCTCCACGATCTAT
	Barcode_579_1859_Rv	TAGATCGTGGAGCGTTGTTCTAGTCTGATAGATCGT
BC09	Barcode_650_1538_Fw	ACGATCTATATGTTGCGAGCACCATTCGACGATCTAT
	Barcode_650_1538_Rv	TAGATCGTCGAATGGTGCTCGCAACATATAGATCGT
BC10	Barcode_749_1984_Fw	ACGATCTACCGTCAGACACTGAACCAGTACGATCTAT
	Barcode_749_1984_Rv	TAGATCGTACTGGTTCAGTGTCTGACGGTAGATCGT
BC11	Barcode_761_1321_Fw	ACGATCTATACGACCACAAGATATTCGGACGATCTAT
	Barcode_761_1321_Rv	TAGATCGTCCGAATATCTTGTGGTCGTATAGATCGT
BC12	Barcode_1183_2117_Fw	ACGATCTAGTCATAACAGGCACTGTCTTACGATCTAT
	Barcode_1183_2117_Rv	TAGATCGTAAGACAGTGCCTGTTATGACTAGATCGT
BC13	Barcode_1438_1303_Fw	ACGATCTACTCACAAGCGATAGTTCCGGACGATCTAT
	Barcode_1438_1303_Rv	TAGATCGTCCGGAACTATCGCTTGTGAGTAGATCGT

BC14	Barcode_1609_1571_Fw	ACGATCTATCCAGTGCTGCGACTACGTGACGATCTAT
	Barcode_1609_1571_Rv	TAGATCGTCACGTAGTCGCAGCACTGGATAGATCGT
BC15	Barcode_1696_1697_Fw	ACGATCTAGATTCGGCACACTCTCGCACACGATCTAT
	Barcode_1696_1697_Rv	TAGATCGTGTGCGAGAGTGTGCCGAATCTAGATCGT
BC16	Barcode_1723_1789_Fw	ACGATCTACAATCCAGGCATCCAGAGCCACGATCTAT
	Barcode_1723_1789_Rv	TAGATCGTGGCTCTGGATGCCTGGATTGTAGATCGT
BC17	Barcode_1835_1863_Fw	ACGATCTAATAGAAGTCCTTATGTCTCCACGATCTAT
	Barcode_1835_1863_Rv	TAGATCGTGGAGACATAAGGACTTCTATTAGATCGT
BC18	Barcode_1946_1947_Fw	ACGATCTACGAGTAGTTCAAGGTAGTTCACGATCTAT
	Barcode_1946_1947_Rv	TAGATCGTGAACTACCTTGAACTACTCGTAGATCGT
BC19	Barcode_2081_2107_Fw	ACGATCTACGCACACGTTCCGCTTGCTTACGATCTAT
	Barcode_2081_2107_Rv	TAGATCGTAAGCAAGCGGAACGTGTGCGTAGATCGT
BC20	Barcode_1975_1976_Fw	ACGATCTAAGCGTTACATTTGACAGCATACGATCTAT
	Barcode_1975_1976_Rv	TAGATCGTATGCTGTCAAATGTAACGCTTAGATCGT

Table S3: Phylogenetic determining variants

Lineage	SNP
1	4357773 GA
2	4357804 TG
3	1281984 GA
4	1281771 CT
5	4357657 GA
6	1281685 CG

Table S4: Percentage of agreement between Illumina and MinION SNPs at different variant calls in MinION and 0.1 frequency in Illumina

Frequency in MinION	TP	FP	FN	TN	TPR	TNR
0.1	140	1478	5	478244	0.9655172	0.996919

0.2	134	93	11	479582	0.9241379	0.9998061
0.3	132	13	14	479654	0.9041096	0.9999729
0.4	**132**	**3**	**14**	**479664**	**0.9041096**	**0.9999937**
0.5	131	2	15	479665	0.8972603	0.9999958
0.6	130	2	16	479665	0.890411	0.9999958
0.7	129	2	17	479665	0.8835616	0.9999958
0.8	125	2	21	479665	0.8561644	0.9999958
0.9	115	2	31	479664	0.7876712	0.9999958

Abbreviations: TP, true positive variants; FN, false negative variants; FP, false positive variants; TN: true negative variants; TPR: true positive rate; TNR: true negative rate.

Table S5: Resistance profile and phylogenetic classification of samples

Sample	Antibiotic							Lineage	Agreement	
	FQ	RMP	SM	INH	PZA	KAN	EMB		Resistance Profile	Lineage
N0067_L1	S	S	S	S	S	S	S	L1	100	100
G981_L2	S	S	S	S	S	S	S	L2	100	100
G107_L3	S	S	S	S	S	S	S	L3	100	100
G770_L4	S	S	S	S	S	S	S	L4	100	100
G1961_L5	S	R	S	S	S	S	R	L5	100	100
G1952_L6	S	S	S	S	S	S	S	L6	100	100
G870	R	R	S	R	S	R	R	L2	100	100
G841	S	R	S	S	S	S	S	L4	100	100
G1800	R	R	S	R	S	R	R	L2	100	100
182320_M20	S	S	S	S	S	S	S	L4	100	100
G1646_W19	S	S	S	S	S	S	S	L4	100	100
G2267_W20	S	S	S	S	S	S	S	L4	100	100
G2103_W23	S	S	S	S	S	S	S	L4	100	100
G2284_W26	S	S	S	S	S	S	S	L4	100	100
G1335_W27	S	S	S	S	S	S	S	L2	100	100

Abbreviations: FQ, fluoroquinolones; RMP, rifampicin; SM, streptomicin; INH, isoniazid; PZA, pyrazinamide; KAN, kanamycin; EMB, ethambutol; L, lineage; S, susceptible; R, resistant.

Table S6: Sequences of reverse primers for the qPCR of the PCR product

Primer	Sequence
katG_Rv_qPCR	CGTGGACCTGGTCTTCGGGTC
gyrA_Rv_qPCR	GGCGGAAGCCGGAATCGAAC
inhA_Rv_qPCR	GCCCTGGCTGCGGGTGTATT
pncA_Rv_qPCR	GGGTGTGCTGCCGATGACGA
gyrB_Rv_qPCR	ATCGATGTGCACCCGCCTGG
embB_Rv_qPCR	CAATCAGCCCGGCGATGGTG
rpoB_Rv_qPCR	GGTCTGGACGTCAAGGAGTCCC
L125_Rv_qPCR	CACCCCGGTGATCCACAGCA
L346_Rv_qPCR	GCGCCACGCCGACATATTCC
eis_Rv_qPCR	TCGCGGTGCTGGTGACGG

rrs_Rv_qPCR	GTCCGAGCGTCTGCACCGAG

Table S7: ROC analysis of vSNPs in MinION (Illumina frequency >= 0.1)

Freq Minion	TP	FP	FN	TN	TPR/Recall	TNR	Sample Type	Agreement
0.1	39	776	0	150696	1	0.9948769	Sputum	4.785276074
0.2	36	48	3	151408	0.9230769	0.9996831	Sputum	42.85714286
0.3	35	4	4	151449	0.8974359	0.9999736	Sputum	89.74358974
0.4	35	3	4	151450	0.8974359	0.9999802	Sputum	92.10526316
0.5	35	2	4	151451	0.8974359	0.9999868	Sputum	94.59459459
0.6	35	2	4	151451	0.8974359	0.9999868	Sputum	94.59459459
0.7	35	2	4	151451	0.8974359	0.9999868	Sputum	94.59459459
0.8	35	2	4	151451	0.8974359	0.9999868	Sputum	94.59459459
0.9	33	2	6	151451	0.8461538	0.9999868	Sputum	94.28571429
0.1	101	702	5	327548	0.9528302	0.9978614	Culture	12.57783313
0.2	98	45	8	328174	0.9245283	0.9998629	Culture	68.53146853
0.3	97	9	10	328205	0.9065421	0.9999726	Culture	91.50943396
0.4	97	0	10	328214	0.9065421	1	Culture	100
0.5	96	0	11	328214	0.8971963	1	Culture	100
0.6	95	0	12	328214	0.8878505	1	Culture	100
0.7	94	0	13	328214	0.8785047	1	Culture	100
0.8	90	0	17	328214	0.8411215	1	Culture	100
0.9	82	0	25	328213	0.7663551	1	Culture	100

Abbreviations: TP, true positive variants; FN, false negative variants; FP, false positive variants; TN: true negative variants; TPR: true positive rate; TNR: true negative rate.

Table S8: ROC analysis of fSNPs in MinION (Illumina frequency >= 0.9)

Freq Minion	TP	FP	FN	TN	TPR/Recall	TNR	Sample Type	Agreement
0.3	35	4	0	151452	1	0.9999736	Sputum	89.74358974
0.4	35	3	0	151453	1	0.9999802	Sputum	92.10526316
0.5	35	2	0	151454	1	0.9999868	Sputum	94.59459459
0.6	35	2	0	151454	1	0.9999868	Sputum	94.59459459
0.7	35	2	0	151454	1	0.9999868	Sputum	94.59459459
0.8	35	2	0	151454	1	0.9999868	Sputum	94.59459459
0.9	33	2	2	151454	0.9428571	0.9999868	Sputum	94.28571429
0.3	94	12	0	328214	1	0.9999634	Culture	88.67924528
0.4	94	3	0	328223	1	0.9999909	Culture	96.90721649
0.5	94	2	0	328224	1	0.9999939	Culture	97.91666667
0.6	94	1	0	328225	1	0.999997	Culture	98.94736842
0.7	94	0	0	328226	1	1	Culture	100
0.8	90	0	4	328226	0.9574468	1	Culture	100
0.9	82	0	12	328225	0.8723404	1	Culture	100

Abbreviations: TP, true positive variants; FN, false negative variants; FP, false positive variants; TN: true negative variants; TPR: true positive rate; TNR: true negative rate.

Table S9: Recall/TNR of drug-resistance and lineage prediction by gene

Prediction	**FQ**		**RMP**	**SM**	**INH**		**PZA**	**KAN**	**EMB**	**L125**	**L346**	**Median**
Gene	***gyrB***	***gyrA***	***rpoB***	***rrs***	***inhA***	***katG***	***pncA***	***eis***	***embB***			
Total SNPs	45675	38115	46330	38867	41445	37110	46138	44989	51330	42355	47460	44989
TP	11	21	16	3	5	11	9	5	19	23	9	11
TN	45664	38094	46310	38863	41440	37099	46129	44983	51311	42316	47448	44983
FP	0	0	0	1	0	0	0	1	0	4	0	0
FN	0	0	4	0	0	0	0	0	0	12	3	0
Recall all SNPs (95% CI)	1 (0.72-1)	1 (0.84-1)	0.8 (0.56-0.94)	1 (0.29-1)	1 (0.48-1)	1 (0.72-1)	1 (0.66-1)	1 (0.78-1)	1 (0.82-1)	0.66 (0.48-0.81)	0.75 (0.43-0.95)	1 (0.72-1)
TNR all SNPs (95% CI)	1 (0.99-1)	1 (0.99-1)	1 (0.99-1)	0.99 (0.99-1)	1 (0.99-1)	1 (0.99-1)	1 (0.99-1)	0.99 (0.99-1)	1 (0.99-1)	0.99 (0.98-1)	1 (0.99-1)	1 (0.99-1)
Agreement all SNPs (%)	100	100	80	75	100	100	100	83.33	100	58.97	75	100
Accuracy all SNPs (%) (95% CI)	100 (99.99-100)	100 (99.99-100)	99.99 (99.99-100)	100 (99.99-100)	100 (99.99-100)	100 (99.99-100)	100 (99.99-100)	100 (99.99-100)	100 (99.99-100)	99.96 (99.94-99.98)	99.99 (99.98-1)	100 (99.99-100)
Sensitivity DR SNPs (95% CI)	1 (0.025-1)	1 (0.025-1)	1 (0.40-1)	No DR SNPs found	No DR SNPs found	1 (0.16-1)	1 (0.025-1)	1 (0.16-1)	1 (0.29-1)	1 (0.54-1)	1 (0.66-1)	1 (0.16-1)
Specificity DR SNPs (95% CI)	1 (0.98-1)	1 (0.97-1)	1 (0.99-1)	1 (0.97-1)	1 (0.88-1)	1 (0.97-1)	1 (0.99-1)	1 (0.94-1)	1 (0.99-1)	1 (0.91-1)	1 (0.90-1)	1 (0.97-1)
Agreement DR SNPs (%)	100	100	100	100	100	100	100	100	100	100	100	100
Accuracy DR SNPs (%) (95% CI)	100 (98.13-100)	100 (96.55-100)	100 (98.89-100)	NA	NA	100 (97.30-100)	100 (99.84-100)	100 (94.04-100)	100 (98.84-100)	100 (92.13-100)	100 (92.13-100)	100 (97-100)

Abbreviations: FQ, fluoroquinolones; RMP, rifampicin; SM, streptomicin; INH, isoniazid; PZA, pyrazinamide; KAN, kanamycin; EMB, ethambutol; L125 and L346, regions containing phylogenetic determining SNPs; CI: confidence intervals; TNP: true negative rate.

SUPPLEMENTARY FIGURES:

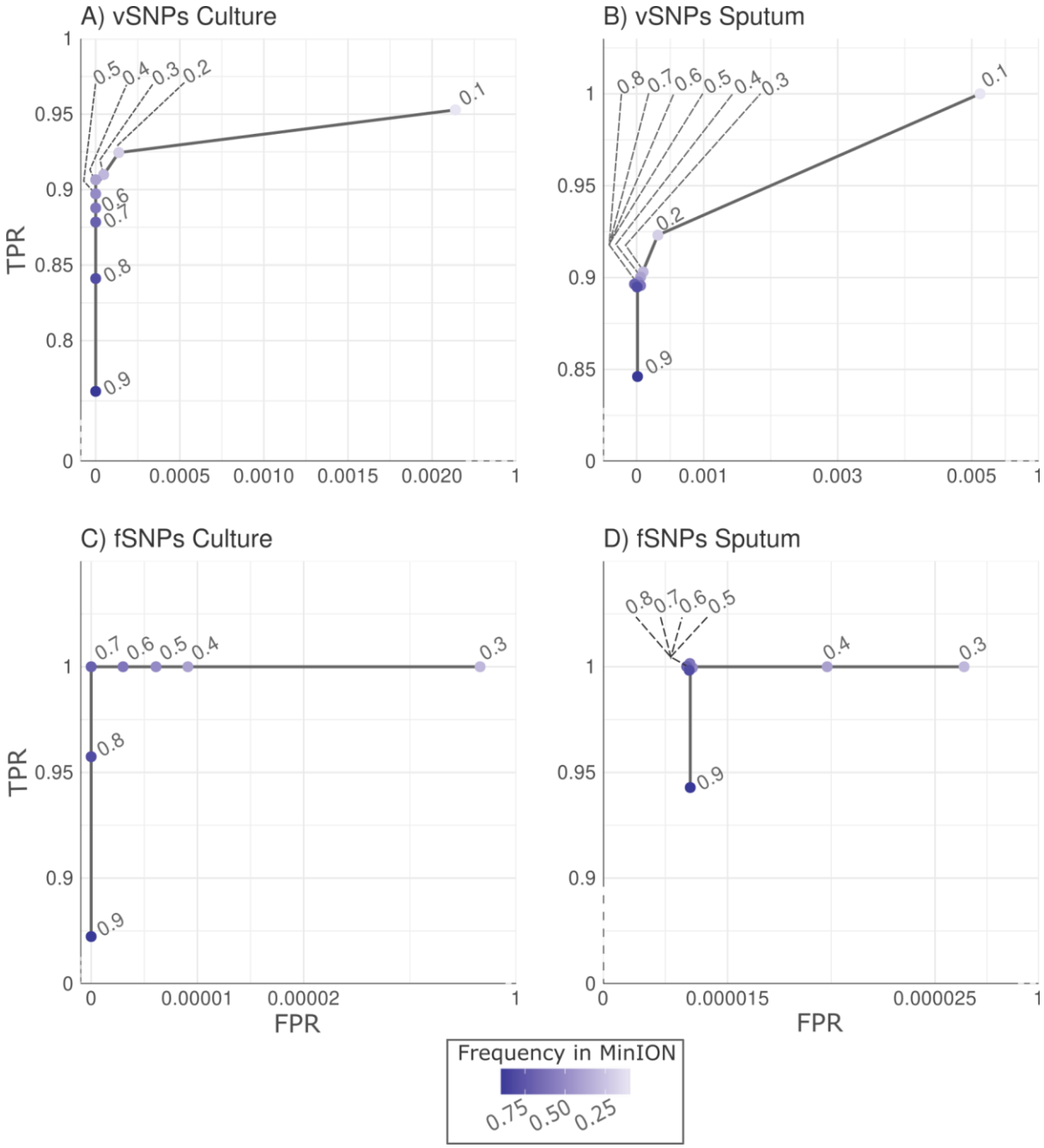

Figure S1: ROC curve used to set the frequency threshold employed to call variants in MinION**.** Points represent the values for recall and false positive rate obtained when applying different frequency values in MinION variant calling. Both axes are truncated. **A-B**) ROC curve used to set the frequency threshold to call variable SNPs in MinION. Recall and false positive rate value obtained using different variant calling frequency cut-offs for MinION (from 0.1 to 0.9 using increments of 0.1) and comparing with Illumina variant calls at a 0.1 fixed threshold. **C-D)** ROC curve used to determine the frequency threshold to call fixed SNPs in MinION. Both axes are truncated. Recall and false positive rate value obtained using different variant calling frequency cut-offs for MinION (from 0.3 to 0.9 using increments of 0.1) and comparing with

Illumina variant calls at a 0.9 fixed threshold. A and C represent the analysis made in culture samples, and B and D represent sputum samples.

SUPPLEMENTARY REFERENCES

1. Buschmann T. DNABarcodes: an R package for the systematic construction of DNA sample tags. Bioinformatics [Internet]. 2017;btw759. Available from: http://dx.doi.org/10.1093/bioinformatics/btw759

2. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol [Internet]. 2019 Jun 24;20(1):129. Available from: http://dx.doi.org/10.1186/s13059-019-1727-y

3. Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved [Internet]. Vol. 42, Nature Genetics. 2010. p. 498–503. Available from: http://dx.doi.org/10.1038/ng.590

4. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics [Internet]. 2018 Sep 15;34(18):3094–100. Available from: http://dx.doi.org/10.1093/bioinformatics/bty191

5. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics [Internet]. 2009 Aug 15;25(16):2078–9. Available from: http://dx.doi.org/10.1093/bioinformatics/btp352

6. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res [Internet]. 2012 Mar;22(3):568–76. Available from: http://dx.doi.org/10.1101/gr.129684.111

7. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Contin Educ Anaesth Crit Care Pain [Internet]. 2008 Dec 1 [cited 2021 May 24];8(6):221–3. Available from: https://academic.oup.com/bjaed/article-pdf/8/6/221/1134124/mkn041.pdf

8. Bainomugisa A, Duarte T, Lavu E, Pandey S, Coulter C, Marais BJ, et al. A complete high-quality MinION nanopore assembly of an extensively drug-resistant Mycobacterium tuberculosis Beijing lineage strain identifies novel variation in repetitive PE/PPE gene regions. Microb Genom [Internet]. 2018 Jul;4(7). Available from: http://dx.doi.org/10.1099/mgen.0.000188