1	A rapid pharmacogenomic assay to detect NAT2 polymorphisms and guide isoniazid dosing
2	for tuberculosis treatment
3	
4	Authors: Renu Verma ¹ , Sunita Patil ¹ , Nan Zhang ² , Flora Martinez Figueira Moreira ³ , Marize
5	Teixeira Vitorio ³ , Andrea da Silva Santos ³ , Ellen Wallace ⁴ , Devasena Gnanashanmugam ⁴ , David
6	Persing ⁴ , Rada Savic ² , Julio Croda ^{5,6} , Jason R. Andrews ^{1*}
7	
8	1. Division of Infectious Diseases and Geographic Medicine, Stanford University School of
9	Medicine, Stanford, CA, USA
10	2. Department of Bioengineering and Therapeutic Sciences, University of California, San
11	Francisco, CA, USA
12	3. Federal University of Grande Dourados, Dourados, Brazil
13	4. Cepheid Inc., Sunnyvale, California, USA
14	5. Postgraduate Program in Infectious and Parasitic Diseases, Federal University of Mato
15	Grosso do Sul, Mato Grosso do Sul, Brazil
16	6. Oswaldo Cruz Foundation Mato Grosso do Sul, Mato Grosso do Sul, Brazil
17	
18	Keywords: tuberculosis; isoniazid; pharmacogenomic; molecular diagnostic; NAT2
19	Abstract Word Count: 183
20	Word Count: 3,817
21	Tables: 3
22	Figures: 4
23	Reference: 50

24

25 Correspondence*

- 26 Jason Andrews, MD
- 27 Division of Infectious Diseases and Geographic Medicine
- 28 Biomedical Innovations Building, Room 3458
- 29 Stanford University School of Medicine
- 30 Stanford, CA 94305
- 31 Email: jandr@stanford.edu
- 32 Phone: +1 650 497 2679

34

35 Abstract

36 Treatment of tuberculosis involves use of standardized weight-based doses of antibiotics, but 37 there remains a substantial incidence of toxicities, inadequate treatment response, and relapse, in 38 part due to variable drug levels achieved. Single nucleotide polymorphisms (SNPs) in the N-39 acetyltransferase-2 (NAT2) gene explain 88% of interindividual pharmacokinetic variability of 40 isoniazid, one of the two most important antitubercular antibiotics. A major obstacle to 41 implementing pharmacogenomic guided dosing is the lack of a point-of-care assay. We trained 42 an acetylation genotype classification model from a global dataset of 5,738 genomes, which 43 achieved 100% accuracy in out-of-sample prediction on unphased SNPs from 2,823 samples 44 using 5 SNPs. On a clinical dataset of 49 patients with tuberculosis, we found that a 5 SNP assay 45 accurately predicted acetylation ratios and isoniazid clearance. We then developed a cartridge-46 based molecular assay for the 5 SNPs on the GeneXpert platform, which enabled accurate 47 classification of allele patterns directly from as little as 25 ul of whole blood. An automated 48 pharmacogenomic assay on a platform widely used globally for tuberculosis diagnosis could 49 enable improved dosing of isoniazid, averting toxicities and improving treatment outcomes.

51 **INTRODUCTION**

52 Despite the availability of effective chemotherapeutic regimens for treatment and prevention of 53 tuberculosis, a substantial proportion of patients experience toxicities, fail treatment or develop recurrent disease¹⁻³. Standardized, weight-based dosing of anti-tuberculosis treatment has been 54 55 the conventional approach to therapy, despite mounting evidence that inter-individual variability in metabolism leads to highly variable drug levels^{4,5}. High drug levels are strongly associated 56 57 with risk of toxicity, while low drug levels are a determinant of treatment failure, slow response, 58 and emergence of drug resistance. Hepatotoxicity is the most common adverse effect, affecting 59 up to 33% of patients receiving standard four-drug therapy⁶ and leading to regimen changes in 60 up to 10% of patients⁷. This toxicity is associated with increased costs, morbidity, and occasional mortality, particularly among HIV co-infected individuals⁸. Additionally, as many as 3% of new 61 tuberculosis cases experience treatment failure, and between 6-10% relapse within 2 years^{9,10} 62 63 Pharmacokinetic variability to a single drug is associated with treatment failure and acquired drug resistance^{11,12}. One recent study found that individuals with at least one drug below the 64 recommended AUC threshold had a 14-fold increased risk of poor outcomes.¹³ 65

There has been an increasing number of genetic markers identified that predict metabolism and toxicities from various antimicrobials. Isoniazid (INH) is among the most well characterized of these, with 88% of its pharmacokinetic variability explained by mutations in the gene encoding arylamine N-acetyltransferase 2 (NAT2), responsible for acetylation in the liver¹⁴. Individuals can be classified into three phenotypes—rapid, intermediate, and slow acetylators—according to whether they carry polymorphisms on neither, one, or both copies of this gene, respectively. Rapid acetylators typically have the lowest plasma INH concentrations, while slow acetylators

73 have high concentrations. A worldwide population survey on NAT2 acetylation phenotype 74 reported that more than half of the global population are slow or rapid acetylators¹⁵. Numerous 75 studies have investigated the relationship between acetylation genotype or phenotype and clinical 76 outcomes of tuberculosis treatment. A recent meta-analysis found that rapid acetylators are twice as likely to have microbiological failure and acquired drug resistance¹⁶. Additional meta-analyses 77 have identified a three- to four-fold increased risk of hepatotoxicity among slow acetvlators^{17,18}. 78 79 A recent randomized trial of pharmacogenomic guided dosing for tuberculosis treatment found 80 that, compared with standard dosing, it reduced hepatotoxicity among slow acetylators and increased treatment response at 8 weeks among rapid acetylators¹⁹ 81

82

83 Despite this evidence, pharmacogenomic testing and guided treatment has not entered the 84 mainstream of clinical practice for tuberculosis. Few clinical laboratories perform NAT2 85 genotyping, and such testing is not widely available in resource-constrained environments where 86 the majority of tuberculosis burden falls. To address this gap, we developed a prototype NAT2 87 pharmacogenomic (NAT2-PGx) assay on a commercial, automated PCR platform (GeneXpert) 88 to detect NAT2 polymorphisms. We further developed an in-house algorithm to predict INH 89 metabolism phenotype from unphased single nucleotide polymorphism (SNP) patterns, derived 90 from globally representative genomic data. We demonstrate that this tool can accurately predict 91 INH clearance rates directly from clinical samples.

92

94 **RESULTS**

95 SNP selection and development of acetylation prediction model

96 Complete phased data for the seven known polymorphisms (191G>A, rs1801279; 282C>T, 97 rs1041983; 341T>C, rs1801280; 481C>T, rs1799929; 590G>A, rs1799930; 803A>G, rs1208; 98 and 857G>A, rs1799931) altering NAT2 gene function were available for 8,561 individuals from 99 59 populations. The dataset contains 3,573 (41.7%) individuals with the slow genotype, 3,428100 (40.0%) individuals with the intermediate genotype, and 1,560 (18.2%) individuals with the 101 rapid genotype (See Table 1). The highest proportion of rapid acetylators were in East Asia (40%), and three regions had prevalence of slow acetylator phenotypes over 50% (Central and 102 103 South Asia, Europe and North Africa).

104 We used these phased allele data to select SNPs for inclusion in an assay measuring unphased 105 SNPs. Using a random forest model trained on two thirds of the data (n=5.738), out-of-sample 106 phenotype prediction accuracy from unphased data on the remaining one third (n=2.823) was 107 100% for models using 7, 6 or 5 SNPs. With 4 SNPs, prediction accuracy was 98.0% (95% CI: 108 97.4-98.5%), and a 3 SNP model had similar performance (98.0%; 95% CI: 97.4-98.4%) (Table 109 2). However, both of these models performed poorly on data from Sub-Saharan Africa (4 SNP 110 model accuracy: 82.5%, 95% CI: 78.1-86.4%); 3 SNP model accuracy: 81.3%, 95% CI: 76.8-111 85.3%). Based on these results, we selected the 5 SNP model (191G>A, 282C>T, 341T>C, 112 590G>A and 857G>A) to take forward for clinical validation and diagnostic development.

113

114 Genotype correlation with isoniazid clearance in patients with tuberculosis

116 We enrolled a cohort of 49 patients with newly diagnosed pulmonary tuberculosis and collected 117 plasma at 1 hour and 8 hours after dose on day 1 and at 1 hour after dose on day 14. To detect 118 five NAT2 polymorphisms identified by our classifier, we performed single-plex PCR assays on 119 49 sputum samples using molecular beacon probes developed in-house (see Methods). 120 Additionally, we used commercial 7-SNP single-plex genotyping assays and compared the 121 results with 5-SNP single-plex PCR to validate the melt curve accuracy. There was 100% 122 concordance in terms of SNP detection between 5-SNP and commercial 7-SNP assays. Of the 49 123 individuals for whom NAT2 genotypes were profiled, ΔTm (°C) between wild-type and mutant 124 alleles for positions 191, 282, 341, 590 and 857 were found to be 4.38, 4.04, 2.40, 3.63 and 3.68 125 respectively. Both mutant and wild type probes had a minimum 2.40°C Tm difference which 126 allowed SNP calling with high accuracy (Table 3).

127

128 We further predicted phenotypes from 5-SNP using the algorithm described above as well as a publicly available tool (NAT2Pred)²⁰, which uses 6 SNPs. Among the 49 participants, predicted 129 acetylator types from the 5 SNP assay were: 28 (57%) slow, 16 (33%) intermediate and 5 (10%) 130 131 rapid. NAT2Pred classified 4 samples as intermediate that were classified as rapid (n=1) or slow 132 (n=3) by the 5 SNP classifier. Among those classified as slow by the 5 SNP classifier and 133 intermediate by NAT2Pred, acetyl-INH to INH ratios at 8 hours were 0.61, 0.38, 0.41, consistent 134 with slow acetylation (median: 0.77, range 0-1.55) rather than intermediate acetylation (median 135 6.67, range 3.32-22.21) and suggesting misclassification by NAT2Pred. The sample classified as 136 intermediate by NAT2PRed and rapid by the 5 SNP classifier had an acetyl-INH to INH ratio of 137 9.8, which fell between the median values, and within both ranges, for intermediate and rapid 138 acetylators (range 8.09 - ∞) (Supplementary Table-1). Phenotypes predicted by the 5 SNP

139 classifier were strongly predictive of INH acetylation and clearance (Figure 1a and 1b). INH 140 clearance rates were lowest in slow acetylators (median 19.3 L/hr), moderate in intermediate 141 acetylators (median 41.0 L/hr) and highest in fast acetylators (median 46.7 L/hr). 142

143 Development of an automated pharmacogenomic assay

144

145 Using the primers and probes sequences validated on FAM-labelled single-plex assays targeting 146 five NAT2 polymorphisms, we developed a 5-plex multiplex assay using the FleXible Cartridge 147 system on the GeneXpert platform, which enables automated extraction, real-time PCR, melt 148 curve analysis and interpretation in 140 minutes (Figure 2). We performed the assay on 20 149 whole blood samples from healthy individuals. Mutant, wild-type and heterozygous alleles were 150 manually called based on peak patterns and Tm values detected in melt curves. Negative 151 derivative transformed melt curves from five NAT2 gene polymorphisms are shown in Figure 3. 152 The genotype data generated on GeneXpert was validated by Sanger sequencing. The assay 153 detected all polymorphisms with 100% accuracy (average SD in Tm across all probes = 0.34° C). 154 The NAT2 genotypes corresponding to 20 blood samples covered all three categories - mutant, 155 wild-type and heterozygous for five NAT2 positions except for NAT2-191 for which all samples 156 were all wild-type. Among the 20 samples, predicted acetylator types using the 5-SNP classifier 157 were: 8 (40%) slow, 10 (50%) intermediate and 2 (10%) rapid (Supplementary Table 2a).

158

159 We used whole blood samples from ten healthy volunteers to assess the analytic performance of 160 the NAT2-PGx assay at lower sample volumes. The samples at decreasing volumes (200ul, 161 100ul, 50ul and 25ul) were analyzed until the point where all melt peaks could be accurately

detected by GeneXpert. Our assay could accurately detect all melt peaks with as low as 25ul of
sample volume. The variability in Tm from five NAT2 probes for sample volumes 200ul-25ul is
shown in Figure 4. NAT2 polymorphisms were accurately detected at all volumes
(Supplementary Table 2b).

167 DISCUSSION

168

169 Despite availability of effective treatment for drug-sensitive tuberculosis, a substantial 170 proportion of population encounters drug associated toxicity or treatment failure, much of which could be averted through dosing guided by genetic markers of drug metabolism^{21,22}. We 171 172 previously found that pharmacogenomic guided dosing of isoniazid could be highly costeffective in low- and middle-income countries²³. A major barrier to its implementation has been 173 174 the lack of a simple, scalable assay that could be used at points of care where tuberculosis is 175 treated in resource-constrained settings. To address this gap, we used globally representative 176 genomic data to identify patterns of 5 SNPs that enable accurate prediction of isoniazid 177 acetylator phenotype, validating this with pharmacokinetic data of patients receiving tuberculosis 178 treatment. We then developed a prototype automated pharmacogenomic assay on the GeneXpert 179 platform, which is widely available globally but had never been applied to pharmacogenomics. 180 We found that this assay could robustly distinguish wild type, mutant and heterozygous alleles 181 from a range of blood volumes as low as 25 ul, making it suitable for use with venous blood 182 samples or finger-stick blood samples. The assay requires minimal hands-on time for sample 183 preparation, which would facilitate its use in resource-constrained settings.

184

An earlier model ("NAT2Pred") predicted NAT2 acetylation phenotype from unphased genomic data; however, it had moderate error rates in distinguishing intermediate from rapid acetylators²⁴. Moreover, error rates among individuals from Sub-Saharan Africa were 14%, in part due to the exclusion of the G191A (R64Q) SNP, common to the NAT2 *14 allele cluster, which is frequent in Africans and African-Americans, but virtually absent in Caucasian, Indian, and Korean

 $populations^{25}$. 190 Indeed. striking ethnic differences in the frequencies of **SNPs** 191 (http://snp500cancer.nci.nih.gov) are responsible for the differences in frequency of rapid, intermediate and slow acetylator *NAT2* alleles or haplotypes and therefore phenotypes^{15, 26, 27}. We 192 193 trained our SNP classifier with globally representative data, which resulted in the selection and 194 inclusion of the G191A SNP in our model and assay. This is particularly important as Sub-195 Saharan Africa bears a substantial burden of tuberculosis disease and mortality as well as HIV 196 co-infection, which is independently associated with greater pharmacokinetic variability and tuberculosis treatment toxicity²⁸⁻³¹. 197

198

199 The association between acetylation polymorphisms and isoniazid metabolism was first 200 demonstrated in 1959, and their importance was well characterized in subsequent decades through phenotypic descriptions³²⁻³⁴. Subsequent genotypic descriptions confirmed that NAT2 201 202 polymorphisms predicted isoniazid early bactericidal activity, and clinical outcomes including 203 hepatotoxicity, relapse and acquisition of drug resistance. Further dosing studies demonstrated 204 that provision of lower doses to slow acetylators and higher doses to rapid acetylators could 205 achieve target concentrations³⁵. One randomized trial of pharmacogenomic-guided dosing of 206 isoniazid during active tuberculosis treatment found that it significantly reduced toxicities 207 (among slow acetylators) and treatment non-response (among rapid acetylators). Taken together, 208 the evidence for pharmacogenomic guided dosing to achieve consistent drug levels and improve 209 clinical outcomes is strong. Automated, easy-to-use assays could enable pharmacogenomic 210 guided isoniazid dosing in resource constrained settings, where a substantial burden of the 211 world's tuberculosis occurs.

212

213 The findings of this study are subject to several limitations. We tested the assays on 49 214 individuals with active tuberculosis and 20 healthy individuals with a diverse representation of 215 polymorphisms, but the number of participants with G191A mutations was limited (n=5). A 216 larger validation study involving testing on whole blood, including from finger stick capillary 217 blood, is needed to assess real-world performance of this assay under field conditions. Further 218 studies should also investigate testing on non-invasive samples including saliva or oral swabs, 219 from which DNA is abundant. Second, we focused on NAT2 polymorphisms, as they explain 220 88% of interindividual pharmacokinetic variability, though polymorphisms in several other 221 genes have been associated with hepatotoxicity. However, these associations have been comparatively modest and somewhat inconsistent³⁶⁻⁴⁰. We focused on isoniazid and did not 222 223 include other important tuberculosis drugs, such as rifampicin. The evidence base for 224 pharmacogenomic markers predicting rifampin pharmacokinetics is less robust, and findings concerning clinical outcomes such as toxicities or treatment response are limited^{41,42}. However, 225 226 given the importance of this drug class in treatment of active and latent tuberculosis, and 227 emerging evidence supporting greater efficacy of higher doses of rifampin, further investigation 228 of pharmacogenomic markers in rifampicin is needed. Future assays may include polymorphisms 229 influencing rifampicin metabolism to further optimize treatment of tuberculosis.

230

Since the demonstration of the efficacy of six month, short-course chemotherapy in 1979, standardized treatment for drug susceptible tuberculosis using weight-based doses has remained essentially unchanged⁴³. Additionally, isoniazid remains a major component of regimens for treatment of latent tuberculosis, which is recommended by the WHO for young children, HIVinfected individuals and household contacts of tuberculosis cases⁴⁴. More than half the world's

236 population have slow or rapid acetylation phenotypes, which put them at risk for excessive drug 237 levels resulting in drug toxicities or insufficient drug levels putting them at risk of acquired drug 238 resistance or disease relapse. Dose adjustment based on NAT2 acetylation genotyping can 239 achieve consistent, target drug levels and reduce the incidence of poor clinical outcomes. We 240 developed a prototype automated, cartridge-based assay that can reliably predict acetylation 241 phenotype directly from as low as 25 ul of whole blood. By developing this for the GeneXpert 242 platform, which is widely used in low- and middle-income countries for tuberculosis diagnosis, 243 this assay could make personalized tuberculosis treatment dosing available in resource-244 constrained settings. Further studies are needed to evaluate its accuracy and clinical impact in 245 real-world clinical settings.

246

247 METHODS

248

249 **Ethics statement**

The clinical study was approved by the institutional review boards of the Stanford University School of Medicine and Federal University of Grande Dourados (IRB#33005). All participants were over the age of 18 and provided written informed consent. For assay optimization and validation on GeneXpert cartridge, anonymized blood samples from healthy individuals were obtained from Stanford blood center.

²⁵⁶ Datasets

The datasets used to develop the NAT2 classifier was obtained from the IGSR (International Genome Sample Resource, 1000 genomes project) and a meta-analysis by Sabbagh et. al^{4,25}. Population information on the combined dataset is provided in **Supplementary Table 3**.

260

NAT2 acetylator phenotype prediction classifier

261 Phased genomes from 8,561 individuals were used and haplotypes were labeled based on 7 262 polymorphic sites in the NAT2 gene (191G>A, 282C>T, 341T>C, 481C>T, 590G>A, 803A>G 263 and 857G>A), following an international consensus nomenclature (43). Individuals with two 264 slow haplotypes were considered slow acetylators; those with two rapid haplotypes were 265 considered rapid acetylators; and those with one slow and one rapid acetylator were considered 266 intermediate acetylators. We then constructed an unphased dataset containing only information 267 on whether each sample was wild type for both alleles, homozygous variant for both alleles, or 268 heterozygous. We then split the dataset into two-thirds for a training set and one third for an out-269 of-sample test set, using sampling stratified by geographic representation to ensure 270 representativeness in the training and test sets. We trained a random forest model on the training set using the *caret* package in R⁴⁵ and assessed classification performance on the test set. We 271 272 began with a 7 SNP model and eliminated SNPs in sequential models according to the lowest 273 variable importance factor.

274

275 Sample collection

276 Sputum and plasma samples from 49 newly diagnosed patients with active pulmonary 277 tuberculosis were collected at the Federal University of Grande Dourados, Brazil. All 278 participants were treated with standardized, weight-based doses of isoniazid, rifampicin,

pyrazinamide and ethambutol. Plasma samples were collected at 1 hour and 8 hours after the first
dose and after 1 hour on day 14. Plasma drug concentrations for isoniazid and acetyl-isoniazid
were quantified by high-performance liquid chromatography coupled to tandem mass
spectrometry (HPLC-MS) as previously described.⁴⁶

283

284 Reference NAT2 SNP genotyping assays

285 Sputum processing, host DNA extraction and single-plex assays on clinical samples

286 Spontaneously expectorated sputum from confirmed pulmonary tuberculosis patients was 287 collected in approximately 10mL of guanidine thiocyanate (GTC) solution (5 M guanidinium 288 thiocyanate, 0.5% w/v sodium N-lauryl sarcosine, 25 mM trisodium citrate, 0.1 M 2mercaptoethanol, 0.5% w/v Tween 80 [pH 7.0]) as described previously⁴⁷. The samples were 289 290 needle sheared and centrifuged at 3000 rpm for 30 min. The supernatant was removed leaving 291 behind 1ml pellet. The pellet was centrifuged at 11,500 rpm for 3 min. Approximately 0.5 ml 292 supernatant was transferred to a fresh cryovial and 0.75 ml Trizol LS was added to the 293 supernatant. The samples were frozen at -80 until used. Host DNA was extracted from the 294 supernatant using a manual extraction method described previously⁴⁸. The DNA was eluted in 50 295 ul DNase-free water and quantified on Qubit. Approximately 10 ng of genomic DNA was used 296 for each single-plex melt curve and commercial TaqMan SNP (NAT2 TaqMan® SNP 297 Genotyping Assays, Applied Biosystems) assay.

298

299 DNA extraction from whole blood samples from healthy individuals

For 5-plex assay validation, genomic DNA from 100ul of whole blood from healthy individuals
was extracted using Qiagen Blood and tissue kit (# 69504). The DNA was eluted in 30ul DNasefree water and quantified on Qubit. 100ng of DNA was used for Sanger sequencing validation.

304 PCR amplification for Sanger DNA sequencing

305 For DNA sequencing, an 823 bp fragment of the NAT2 gene (819–1641 bp of the gene) was 306 amplified using the forward primer 5'-GGGCTGTTCCCTTTGAGA-3' and reverse primer 5'-307 TAGTGAGTTGGGTGATAC-3'. A 20 µl PCR mixture contained 0.5 µl of each forward and 308 reverse primers from 10 µM stocks, 8 µl double distilled water, 0.5ul Phusion Tag polymerase 309 and 1 μ l (~100ng) DNA template. PCR was performed with initial denaturation at 95°C for 5 310 min followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s and 311 extension at 72°C for 1 min, with an additional extension at 72°C for 10 min. PCR products were 312 analyzed on 1.5% agarose gels to confirm size of product which was then sequenced at the 313 (Stanford PAN Facility, CA).

314

315 **Primers and probes for melt curve analysis**

Three sets of primers spanning the *NAT2* gene were used for single-plex and multiplex PCR. Primers and sloppy molecular beacon (SMB) probes were designed using Beacon Designer (Premier Biosoft International, Palo, CA; version 8). Three of 5 molecular beacon probes (NAT2-191, NAT2-590 and NAT2-857) were designed with 100% complementarity towards mutant alleles and two (NAT2-282 and NAT2-341) were 100% specific to wild type alleles. For single-plex assays, all probes were labelled with FAM at their 5' end and BHQ-1 at 3'. For the multiplex assay on GeneXpert, FAM was replaced with other fluorophores except for NAT2-

590. A list of primers and probes sequences for the multiplex assay with their corresponding
fluorophores and quenchers is provided in **Supplementary Table 4.**

325

326 Single-plex PCR and melt curve analysis on pulmonary TB patients

327 Genomic DNA extracted from sputum samples from TB positive patients was used for single-328 plex assays performed on StepOne Plus Real Time PCR. A 20ul total reaction volume was set up 329 using 10ng genomic DNA per assay. PCR mastermix included (0.5 ul of 2U Aptatag exo-DNA 330 polymerase, 1X betaine, 1X Aptataq buffer, 4mM MgCl2, 1X ROX passive reference dve, 60nM 331 FP,1000nM reverse primers, 250uM of each probe). PCR was initiated by 10 min of 332 denaturation–activation at 95°C, followed by 50 cycles at 95°C for 15 sec (denaturation), 333 annealing at 60°C for 15 s and extension at 76°C for 20sec. The melting program included three 334 steps: denaturation at 95°C for 1 min, followed by renaturation at 35°C for 3 min and a 335 continuous reading of fluorescence from 45 to 85°C by heating at increments of 0.03°C/s. The 336 MMCA curve was analyzed using the StepOne Plus software version 2.0. For single-plex melt 337 curve assay validation, TaqMan 7-SNP genotyping assays were performed using commercial 338 assays (NAT2 TaqMan® SNP Genotyping Assays, Applied Biosystems) on DNA extracted from 339 sputum samples from 49 pulmonary TB patients on a StepOne Plus Real Time PCR machine.

340

341 Pharmacokinetic analysis of INH clearance in tuberculosis patients

The population PK analysis was performed using the non-linear mixed effects modeling approach using NONMEM (version 7.4.3; ICON plc, Gaithersburg, MD, USA), PsN and Rbased Xpose (version 4.7 and higher)^{49,50}. One-compartment model with a first-order absorption with a lognormal distribution for inter-individual variability (IIV) on different PK parameter(s)

346 as well as an additive and/or proportional model for the residual error were tested for the base 347 model selection. Mixture models with two or three subpopulations representing different 348 clearance rate were then evaluated. The first-order conditional estimation with interaction 349 method (FOCEI) was applied and the model-building procedure and model selection was based 350 on the log-likelihood criterion (the difference in the minimum OFV between hierarchical models 351 was assumed to be Chi-square distributed with degrees of freedom equal to the difference in the 352 number of parameters between models), goodness-of-fit plots (e.g. relevant residuals against 353 time randomly distributed around zero), and scientific plausibility of the model. Visual predictive 354 check was conducted to evaluate whether the final model with estimated fixed-effect parameters 355 and covariates adequately describe data.

356

357 Automated NAT2-PGx Multiplex PCR and melt curve analysis

358 Asymmetric PCR and melt curve analysis were performed on a GeneXpert IV instrument using 359 GeneXpert Dx 4.8 software (Cepheid, Sunnyvale). Flex cartridge-01 (Cepheid) were used to 360 perform automated DNA extraction from whole blood followed by PCR amplification and melt 361 curve analysis to detect SNPs. PCR and melt conditions were optimized using mastermix 362 prepared in house. The NAT2-PGx assay was performed in an 80ul reaction volume (70ul 363 mastermix and 10ul eluted DNA). Briefly, 100ul of whole blood was mixed with 900ul of lysis 364 buffer (Cepheid) in a 1.5ml Eppendorf tube. The sample was vortexed for 2-3 sec and incubated 365 at room temperature for 2min. The entire 1ml whole blood and lysis buffer mix was loaded into 366 sample preparation chamber of flex cart-01 for automated DNA extraction. 70ul PCR masternix 367 was simultaneously loaded in the PCR reaction chamber of the flex cart-01. The GeneXpert was programmed to elute DNA in 10ul volume which was used for the NAT2-PGx assay. PCR 368

369 mastermix included (2ul of 2U Aptataq exo-DNA polymerase, 1X Betaine, 1X PCR additive 370 reagent, 1X Aptataq buffer, 8mM MgCl2, 400nM FP, 900nM reverse primers, 500nM of 191-371 Cv5.5 and 857-Alexa-405, 430nM 282-Alexa-647 and 590-FAM and 300nM 341-Alexa-537 372 probes). PCR was initiated at holding stage- 50°C for 2min, initial denaturation at 94°C for 373 2min, followed by 50 cycles of denaturation at 95°C for 15 sec, annealing at 57°C for 30 sec and 374 extension at 65°C for 60 sec. The melting program included three steps: denaturation at 95°C for 375 1 min, followed by renaturation at 40°C for 3 min and a continuous reading of fluorescence from 376 40 to 72°C by heating at increments of 0.05°C/sec. The MMCA curve was analyzed using the 377 GXP version 4.8 software.

378

We validated the automated NAT2-PGx assay by analyzing 20 blood samples, for which polymorphisms in the 5 positions were confirmed by Sanger sequencing. We assessed analytical sensitivity of the assay and robustness to input blood volume by performing it on varying volumes of whole blood (200 ul, 100 ul, 50 ul, 25 ul) and comparing the Tm results and standard deviation for each position across blood volumes.

384

- 385 Funding
- 386

FleXcartridges and technical support for their use were provided by Cepheid. This study was
supported by the Institute for Immunity, Transplantation, and Infection and the Department of
Medicine at Stanford University.

390

391 Reporting summary. Further information on research design is available in the Nature Research
392 Reporting Summary linked to this article.

393

394 Data availability

- 395 Data supporting the findings of this manuscript are available in the Supplementary Information
- 396 files or from the corresponding author upon request.

397

398 Code availability

- 399
- 400 Code and dataset used to develop NAT2 classifier are provided with the manuscript.

402 Acknowledgements

403 We thank Veronique Dartois for performing the plasma drug level assays.

404

405 Author contributions

- 406 JRA and RV conceived of the study. RV, JC, and JRA designed the experiments. RV, SP,
- 407 FMM, MTV, ASS and JC collected data. RV, NZ, RS and JRA analyzed data. EW, DG and DP
- 408 provided technical guidance on the assay development. RV and JRA wrote the first draft of the
- 409 manuscript, and all authors contributed to the final version.

410 **Conflicts of Interest**

- 411 JRA and RV are named co-inventors on a provisional patent (Application number 62/991,477) for
- 412 a *NAT2* pharmacogenomic assay.

413 **References**

- Gillespie SH. et al. Consortium. Four-month moxifloxacin-based regimens for drug sensitive tuberculosis. N Engl J Med. 2014 Oct 23;371(17):1577-87.
- 416
 2. Merle CS. et al. OFLOTUB/Gatifloxacin for Tuberculosis Project. A four-month
 417 gatifloxacin-containing regimen for treating tuberculosis. N Engl J Med. 2014 Oct
 418 23;371(17):1588-98.
- Jindani A. et al. High-dose rifapentine with moxifloxacin for pulmonary tuberculosis. N
 Engl J Med. 2014 Oct 23:371(17):1599-608.
- 4. Pratt VM. et al. Characterization of 137 Genomic DNA Reference Materials for 28
 Pharmacogenetic Genes: A GeT-RM Collaborative Project. J Mol Diagn. 2016
 Jan;18(1):109-23.
- 424 5. García-Martín E. Interethnic and intraethnic variability of NAT2 single nucleotide
 425 polymorphisms. Curr Drug Metab. 2008 Jul;9(6):487-97.
- 426 6. Schaberg T et al. Risk factors for side-effects of isoniazid, rifampin and pyrazinamide in
 427 patients hospitalized for pulmonary tuberculosis. Eur Respir J. 1996 Oct;9(10):2026-30.
- 428 7. Saukkonen JJ. et al. Hepatotoxicity of Antituberculosis Therapy Subcommittee. An
 429 official ATS statement: hepatotoxicity of antituberculosis therapy. Am J Respir Crit Care
 430 Med. 2006 Oct 15;174(8):935-52.
- 431 8. Schutz C. et al. Burden of antituberculosis and antiretroviral drug-induced liver injury at
 432 a secondary hospital in South Africa. S Afr Med J. 2012 Mar 2;102(6):506-11.
- 433 9. Luzze H. et al. Relapse more common than reinfection in recurrent tuberculosis 1-2 years
- 434 post treatment in urban Uganda. Int J Tuberc Lung Dis. 2013 Mar;17(3):361-7

- 435 10. Romanowski K. et al. Predicting tuberculosis relapse in patients treated with the standard
 436 6-month regimen: an individual patient data meta-analysis. Thorax. 2019 Mar;74(3):291437 297.
- 438 11. Pasipanodya JG. et al. Meta-analysis of clinical studies supports the pharmacokinetic
 439 variability hypothesis for acquired drug resistance and failure of antituberculosis therapy.
- 440 Clin Infect Dis. 2012 Jul;55(2):169-77.
- 441 12. Gumbo T. et al. Isoniazid bactericidal activity and resistance emergence: integrating
 442 pharmacodynamics and pharmacogenomics to predict efficacy in different ethnic
 443 populations. Antimicrob Agents Chemother. 2007 Jul;51(7):2329-36.
- 444 13. Pasipanodya JG. et al. Serum drug concentrations predictive of pulmonary tuberculosis
 445 outcomes. J Infect Dis. 2013 Nov 1;208(9):1464-73.
- 446 14. Kinzig-Schippers M, Tomalik-Scharte D, Jetter A, Scheidel B, Jakob V, Rodamer M,
 447 Cascorbi I, Doroshyenko O, Sörgel F, Fuhr U. Should we use N-acetyltransferase type 2
 448 genotyping to personalize isoniazid doses? Antimicrob Agents Chemother. 2005
 449 May;49(5):1733-8.
- 450 15. Sabbagh A. et al. Worldwide distribution of NAT2 diversity: implications for NAT2
 451 evolutionary history. BMC Genet. 2008 Feb 27;9:21.
- 452 16. Wang PY. et al. NAT2 polymorphisms and susceptibility to anti-tuberculosis drug453 induced liver injury: a meta-analysis. Int J Tuberc Lung Dis. 2012 May;16(5):589-95.
- 454 17. Du H. et al. Slow N-acetyltransferase 2 genotype contributes to anti-tuberculosis drug455 induced hepatotoxicity: a meta-analysis. Mol Biol Rep. 2013 May;40(5):3591-6.
- 456 18. Azuma J. et al. Pharmacogenetics-based tuberculosis therapy research group. NAT2
 457 genotype guided regimen reduces isoniazid-induced liver injury and early treatment

458	failure in the 6-month four-drug standard treatment of tuberculosis: a randomized
459	controlled trial for pharmacogenetics-based therapy. Eur J Clin Pharmacol. 2013
460	May;69(5):1091-101.
461	19. Snider DE Jr. et al. Isoniazid-associated hepatitis deaths: a review of available
462	information. Am Rev Respir Dis. 1992 Feb;145(2 Pt 1):494-7.
463	20. Kuznetsov IB. et al. A web server for inferring the human N-acetyltransferase-2 (NAT2)
464	enzymatic phenotype from NAT2 genotype. Bioinformatics. 2009 May 1;25(9):1185-6.
465	21. Ramachandran G. et al. Role of pharmacogenomics in the treatment of tuberculosis: a
466	review. Pharmgenomics Pers Med. 2012;5:89-98.
467	22. McInnes G. et al. Drug Response Pharmacogenetics for 200,000 UK Biobank
468	Participants. doi: https://doi.org/10.1101/2020.08.09.243311
469	23. Rens NE. et al. Cost-Effectiveness of a Pharmacogenomic Test for Stratified Isoniazid
470	Dosing in Treatment of Active Tuberculosis. Clin Infect Dis. 2020 Jan 6:ciz1212.
471	24. Sabbagh A. et al. Evaluating NAT2PRED for inferring the individual acetylation status
472	from unphased genotype data. BMC Med Genet. 2009 Dec 31;10:148.
473	25. Sabbagh A. et al. Arylamine N-acetyltransferase 2 (NAT2) genetic diversity and
474	traditional subsistence: a worldwide population survey. PLoS One. 2011 Apr
475	6;6(4):e18507.
476	26. Van Oosterhout JJ. et al. Pharmacokinetics of Antituberculosis Drugs in HIV-Positive
477	and HIV-Negative Adults in Malawi. Antimicrob Agents Chemother. 2015
478	Oct;59(10):6175-80.
479	27. Wilkins JJ. et al. Variability in the population pharmacokinetics of isoniazid in South
480	African tuberculosis patients. Br J Clin Pharmacol. 2011 Jul;72(1):51-62.

- 481 28. Wilkins JJ. et al. Population pharmacokinetics of rifampin in pulmonary tuberculosis
- 482 patients, including a semimechanistic model to describe variable absorption. Antimicrob

483 Agents Chemother. 2008 Jun;52(6):2138-48.

- 484 29. Wilkins JJ. et al. Variability in the population pharmacokinetics of pyrazinamide in South
- 485 African tuberculosis patients. Eur J Clin Pharmacol. 2006 Sep;62(9):727-35.
- 486 30. Iselius L. et al. Formal genetics of isoniazid metabolism in man. Clin Pharmacokinet.
 487 1983 Nov-Dec;8(6):541-4.
- 488 31. Evans DA. Genetic variations in the acetylation of isoniazid and other drugs. Ann N Y
 489 Acad Sci. 1968 Jul 31;151(2):723-33.
- 490 32. Evans DA. An improved and simplified method of detecting the acetylator phenotype. J
 491 Med Genet. 1969 Dec;6(4):405-7.
- 33. Donald PR. et al. The influence of dose and N-acetyltransferase-2 (NAT2) genotype and
 phenotype on the pharmacokinetics and pharmacodynamics of isoniazid. Eur J Clin
 Pharmacol. 2007 Jul;63(7):633-9.
- 495 34. Lee SW. et al. NAT2 and CYP2E1 polymorphisms and susceptibility to first-line anti496 tuberculosis drug-induced hepatitis. Int J Tuberc Lung Dis. 2010 May;14(5):622-6.
- 497 35. Roy B. et al. Predisposition of antituberculosis drug induced hepatotoxicity by
 498 cytochrome P450 2E1 genotype and haplotype in pediatric patients. J Gastroenterol
 499 Hepatol. 2006 Apr;21(4):784-6.
- 36. Simon T. et al. Combined glutathione-S-transferase M1 and T1 genetic polymorphism
 and tacrine hepatotoxicity. Clin Pharmacol Ther. 2000 Apr;67(4):432-7.
- 502 37. Strange RC. et al. Glutathione S-transferase: genetics and role in toxicology. Toxicol
 503 Lett. 2000 Mar 15;112-113:357-63.

- 38. Ramachandran G. et al. Role of pharmacogenomics in the treatment of tuberculosis: a
 review. Pharmgenomics Pers Med. 2012;5:89-98.
- 50639. Mukonzo JK. et al. Role of pharmacogenetics in rifampicin pharmacokinetics and the507potential effect on TB-rifampicin sensitivity among Ugandan patients. Trans R Soc Trop
- 508 Med Hyg. 2020 Feb 7;114(2):107-114.
- 40. Naidoo A. et al. Effects of genetic variability on rifampicin and isoniazid
 pharmacokinetics in South African patients with recurrent tuberculosis.
 Pharmacogenomics. 2019 Mar;20(4):225-240.
- 41. Medellin-Garibay SE. et al. A population approach of rifampicin pharmacogenetics and
 pharmacokinetics in Mexican patients with tuberculosis. Tuberculosis (Edinb). 2020
 Sep;124:101982.
- 515 42. *Database of arylamine N-acetyltransferases* (http://nat.mbg.duth.gr)
- 516 43. Clinical trial of six-month and four-month regimens of chemotherapy in the treatment of
 517 pulmonary tuberculosis. Am Rev Respir Dis. 1979 Apr;119(4):579-85.
- 518 44. Haley CA. Treatment of Latent Tuberculosis Infection. Microbiol Spectr. 2017 Apr;5(2).
- 519 45. Kuhn M. et al. (2019). caret: Classification and Regression Training. R package version
 520 6.0-84. https://CRAN.R-project.org/package=caret
- 46. Prideaux B. et al. The association between sterilizing activity and drug distribution into
 tuberculosis lesions. Nat Med. 2015 Oct;21(10):1223
- 47. Walter ND. et al. Transcriptional Adaptation of Drug-tolerant Mycobacterium
 tuberculosis During Treatment of Human Tuberculosis. J Infect Dis. 2015 Sep
 15;212(6):990-8.

526	48. Chomczynski F	P. A reagen	t for the	single-step	simultaneous	isolation	of RNA,	DNA	and

- 527 proteins from cell and tissue samples. Biotechniques. 1993 Sep;15(3):532-4, 536-7.
- 528 49. Beal SL. Et al. NONMEM 7.4 Users Guides. (1989–2019). ICON plc, Gaithersburg,
- 529 MD. https://nonmem.iconplc.com/nonmem744
- 530 50. Lindbom L. et al. (2004). Perl-speaks-NONMEM (PsN)--a Perl module for NONMEM
- 531 related programming. *Comput Methods Programs Biomed*, 75 (2), 85-94.
- 532
- 533
- 534
- 535

Figure 1. Predicted *NAT2* phenotype from sputum samples and associated acetylation ratio and isoniazid clearance rates from patients receiving tuberculosis treatment. The (a) 8 hour acetyl-INH to INH ratio and (b) isonazid clearance rates, according to acetylation phenotype predicted from 5 SNPs, measured in sputum samples from 49 patients receiving treatment for active tuberculosis.

541

Figure 2. Schemata for the automated NAT2 Pharmacogenomic assay. 1-2 drops of blood is collected in an Eppendorf tube and mixed with lysis buffer to a total of 1 ml, which is then loaded onto a GeneXpert Flex01 cartridge and placed into a GeneXpert instrument for automated DNA extraction, asymmetric PCR and meltcurve analysis. Allele patterns for each of the 5 SNPs are determined by Tm analysis, and the resulting data are used to predict acetylator phenotype.

547

Figure 3. Negative derivative transformed melt curves for the five NAT2 gene polymorphisms. The shift in melt curve temperature is observed during a nucleotide exchange. Sloppy molecular beacon probes are first hybridized and then melted off of their NAT2 target amplicon. The melt curves indicate wild type alleles at positions 191(red), 341(green) and 857 (orange); and mutant alleles at positions 282(Blue) and 590(Purple).

553

Figure 4. Effect of whole blood sample volume on melting temperature for wild type and mutant alleles at 5 positions in *NAT2*: *NAT2* polymorphisms were accurately detected at all volumes with sufficient different in melting temperature (Tm) to distinguish wild type from mutant alleles. No individuals in this dataset had mutations at position 191.

558

559 Table 1. Summary of populations included in genomic analysis and their acetylation

560 genotypes.

561

		Acetylation Genotype, n (%)						
	Number of							
Region	individuals	Slow	Intermediate	Rapid				
Americas	1,112	432 (39%)	463 (42%)	217 (20%)				
Central and South Asia	588	355 (60%)	198 (34%)	35 (6%)				
East Asia	2,308	340 (15%)	1049 (45%)	919 (40%)				
Europe	3,458	1966 (57%)	1249 (36%)	243 (7%)				
North Africa	44	30 (68%)	10 (23%)	4 (9%)				
sub-Saharan Africa	1,051	450 (43%)	459 (44%)	142 (14%)				
Total	8,561	3573 (42%)	3428 (40%)	1560 (18%)				

562

564 Table 2. Out-of-sample prediction accuracy of unphased NAT2 SNP data for acetylation

565 phenotype in random forest models. Models were trained with 5,738 individuals and tested on

566 2,823 individuals. Sens: sensitivity. Spec: specificity.

NAT2 SNP Positions	SNP number	Accuracy	95% CI	Sens. Rapid	Spec. Rapid	Sens. Slow	Spec Slow
191, 282, 341, 481, 590, 803, 857	7	1.000	(0.999-1.000)	1.000	1.000	1.000	1.000
191, 282, 341, 481, 590, 857	6	1.000	(0.999-1.000)	1.000	1.000	1.000	1.000
191, 282, 341, 590, 857	5	1.000	(0.999-1.000)	1.000	1.000	1.000	1.000
282, 341, 590, 857	4	0.978	(0.972-0.983)	0.996	0.988	0.969	0.999
341, 590, 847	3	0.976	(0.970-0.982)	1.000	0.986	0.967	1.000
341, 590	2	0.852	(0.838-0.865)	1.000	0.889	0.832	1.000

575 Table 3. Melting temperature (Tm) values for five NAT2 polymorphisms derived from

576 **DNA-probe hybrid melts using single-plex assays validated on 49 pulmonary TB patients**

577

	Mutant	Wild type	Het	Mutant	Wild type	
NAT2 SNP position	Total samples analyzed	Total samples analyzed	Total samples analyzed	$Tm \pm SD$	$Tm \pm SD$	ΔTm (WT- MT)
191	0	44	5	66.3 ± 0.28	$\begin{array}{c} 61.92 \pm \\ 0.68 \end{array}$	4.38
282	6	24	19	62.81 ± 0.40	66.85 ± 0.34	4.04
341	12	23	14	66.30 ± 0.48	$\begin{array}{c} 68.70 \pm \\ 0.38 \end{array}$	2.40
590	0	35	14	68.0 ± 0.28	$\begin{array}{c} 64.37 \pm \\ 0.28 \end{array}$	3.63
857	1	41	7	66.4± 0.11	62.72 ± 0.29	3.68

578 SD: Standard deviation

580

581	Supplementary Table 1. NAT2 polymorphism profiles of 49 pulmonary TB patients detected
582	from 5-SNP and 7-SNP genotyping assays. Acetylator phenotypes predicted from 5-SNP, 6-SNP
583	and 7-SNP classifiers and plasma INH and acetyl-INH levels recorded at various time-points.
584	
585	Supplementary Table 2. (a) NAT2 genotype of whole blood samples (n=20) analyzed on 5-
586	plex NAT2-PGx assay and validated on Sanger sequencing. (b) NAT2 genotypes of ten whole
587	blood samples analyzed using 200ul, 100ul, 50ul and 25ul samples
588	
589	Supplementary Table 3. Population information on the dataset used to develop 5-SNP
590	classifier.
591	
592	Supplementary Table 4. List of primers and probes sequences for multiplex assay with their
593	corresponding fluorophores and quenchers
594	





Sample collection



Collect 1-2 drops of whole blood (approx. 50ul)



Automated DNA extraction & qPCR



Total run

Load the cartridge in the GeneXpert machine and start the run





Mix 950ul lysis buffer, vortex for 2 sec and incubate at RT for 2min

Data aquisition (5)



Acquire melt curve data for NAT2 gene polymorphism detection

Flex cart01 3 preparation



Pipette 1ml pretreated sample and 70ul PCR mastermix into the cartridge

INH acetylator (6) phenotype prediction



Upload NAT2 polymorphism data on phenotype prediction algorithm



