A rapid pharmacogenomic assay to detect \textit{NAT2} polymorphisms and guide isoniazid dosing for tuberculosis treatment

\textbf{Authors:} Renu Verma\textsuperscript{1}, Sunita Patil\textsuperscript{1}, Nan Zhang\textsuperscript{2}, Flora Martinez Figueira Moreira\textsuperscript{3}, Marize Teixeira Vitorio\textsuperscript{3}, Andrea da Silva Santos\textsuperscript{3}, Ellen Wallace\textsuperscript{4}, Devasena Gnanashanmugam\textsuperscript{4}, David Persing\textsuperscript{4}, Rada Savic\textsuperscript{2}, Julio Croda\textsuperscript{5,6}, Jason R. Andrews\textsuperscript{1*}

1. Division of Infectious Diseases and Geographic Medicine, Stanford University School of Medicine, Stanford, CA, USA
2. Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA
3. Federal University of Grande Dourados, Dourados, Brazil
4. Cepheid Inc., Sunnyvale, California, USA
5. Postgraduate Program in Infectious and Parasitic Diseases, Federal University of Mato Grosso do Sul, Mato Grosso do Sul, Brazil
6. Oswaldo Cruz Foundation Mato Grosso do Sul, Mato Grosso do Sul, Brazil

\textbf{Keywords:} tuberculosis; isoniazid; pharmacogenomic; molecular diagnostic; \textit{NAT2}

\textbf{Abstract Word Count:} 183

\textbf{Word Count:} 3,817

\textbf{Tables:} 3

\textbf{Figures:} 4

\textbf{Reference:} 50

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Correspondence*

Jason Andrews, MD
Division of Infectious Diseases and Geographic Medicine
Biomedical Innovations Building, Room 3458
Stanford University School of Medicine
Stanford, CA 94305
Email: jandr@stanford.edu
Phone: +1 650 497 2679
Abstract

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

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

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 liver. 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
have high concentrations. A worldwide population survey on NAT2 acetylation phenotype reported that more than half of the global population are slow or rapid acetylators\textsuperscript{15}. Numerous studies have investigated the relationship between acetylation genotype or phenotype and clinical outcomes of tuberculosis treatment. A recent meta-analysis found that rapid acetylators are twice as likely to have microbiological failure and acquired drug resistance\textsuperscript{16}. Additional meta-analyses have identified a three- to four-fold increased risk of hepatotoxicity among slow acetylators\textsuperscript{17,18}. A recent randomized trial of pharmacogenomic guided dosing for tuberculosis treatment found that, compared with standard dosing, it reduced hepatotoxicity among slow acetylators and increased treatment response at 8 weeks among rapid acetylators\textsuperscript{19}

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

SNP selection and development of acetylation prediction model

Complete phased data for the seven known polymorphisms (191G>A, rs1801279; 282C>T, rs1041983; 341T>C, rs1801280; 481C>T, rs1799929; 590G>A, rs1799930; 803A>G, rs1208; and 857G>A, rs1799931) altering NAT2 gene function were available for 8,561 individuals from 59 populations. The dataset contains 3,573 (41.7%) individuals with the slow genotype, 3,428 (40.0%) individuals with the intermediate genotype, and 1,560 (18.2%) individuals with the 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 South Asia, Europe and North Africa).

We used these phased allele data to select SNPs for inclusion in an assay measuring unphased SNPs. Using a random forest model trained on two thirds of the data (n=5,738), out-of-sample phenotype prediction accuracy from unphased data on the remaining one third (n=2,823) was 100% for models using 7, 6 or 5 SNPs. With 4 SNPs, prediction accuracy was 98.0% (95% CI: 97.4-98.5%), and a 3 SNP model had similar performance (98.0%; 95% CI: 97.4-98.4%) (Table 2). However, both of these models performed poorly on data from Sub-Saharan Africa (4 SNP model accuracy: 82.5%, 95% CI: 78.1-86.4%); 3 SNP model accuracy: 81.3%, 95% CI: 76.8-85.3%). Based on these results, we selected the 5 SNP model (191G>A, 282C>T, 341T>C, 590G>A and 857G>A) to take forward for clinical validation and diagnostic development.

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

We further predicted phenotypes from 5-SNP using the algorithm described above as well as a publicly available tool (NAT2Pred)\textsuperscript{20}, which uses 6 SNPs. Among the 49 participants, predicted acetylator types from the 5 SNP assay were: 28 (57%) slow, 16 (33%) intermediate and 5 (10%) rapid. NAT2Pred classified 4 samples as intermediate that were classified as rapid (n=1) or slow (n=3) by the 5 SNP classifier. Among those classified as slow by the 5 SNP classifier and intermediate by NAT2Pred, acetyl-INH to INH ratios at 8 hours were 0.61, 0.38, 0.41, consistent with slow acetylation (median: 0.77, range 0-1.55) rather than intermediate acetylation (median 6.67, range 3.32-22.21) and suggesting misclassification by NAT2Pred. The sample classified as intermediate by NAT2Pred and rapid by the 5 SNP classifier had an acetyl-INH to INH ratio of 9.8, which fell between the median values, and within both ranges, for intermediate and rapid acetylators (range 8.09 - $\infty$) (Supplementary Table-1). Phenotypes predicted by the 5 SNP
classifier were strongly predictive of INH acetylation and clearance (Figure 1a and 1b). INH clearance rates were lowest in slow acetylators (median 19.3 L/hr), moderate in intermediate acetylators (median 41.0 L/hr) and highest in fast acetylators (median 46.7 L/hr).

**Development of an automated pharmacogenomic assay**

Using the primers and probes sequences validated on FAM-labelled single-plex assays targeting five \textit{NAT2} polymorphisms, we developed a 5-plex multiplex assay using the FleXible Cartridge system on the GeneXpert platform, which enables automated extraction, real-time PCR, melt curve analysis and interpretation in 140 minutes (Figure 2). We performed the assay on 20 whole blood samples from healthy individuals. Mutant, wild-type and heterozygous alleles were manually called based on peak patterns and Tm values detected in melt curves. Negative derivative transformed melt curves from five \textit{NAT2} gene polymorphisms are shown in Figure 3. The genotype data generated on GeneXpert was validated by Sanger sequencing. The assay detected all polymorphisms with 100% accuracy (average SD in Tm across all probes = 0.34°C).

The \textit{NAT2} genotypes corresponding to 20 blood samples covered all three categories - mutant, wild-type and heterozygous for five \textit{NAT2} positions except for \textit{NAT2}-191 for which all samples were all wild-type. Among the 20 samples, predicted acetylator types using the 5-SNP classifier were: 8 (40%) slow, 10 (50%) intermediate and 2 (10%) rapid (Supplementary Table 2a).

We used whole blood samples from ten healthy volunteers to assess the analytic performance of the \textit{NAT2}-PGx assay at lower sample volumes. The samples at decreasing volumes (200ul, 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).
DISCUSSION

Despite availability of effective treatment for drug-sensitive tuberculosis, a substantial 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. We previously found that pharmacogenomic guided dosing of isoniazid could be highly cost-effective in low- and middle-income countries. A major barrier to its implementation has been the lack of a simple, scalable assay that could be used at points of care where tuberculosis is treated in resource-constrained settings. To address this gap, we used globally representative genomic data to identify patterns of 5 SNPs that enable accurate prediction of isoniazid acetylator phenotype, validating this with pharmacokinetic data of patients receiving tuberculosis treatment. We then developed a prototype automated pharmacogenomic assay on the GeneXpert platform, which is widely available globally but had never been applied to pharmacogenomics. We found that this assay could robustly distinguish wild type, mutant and heterozygous alleles from a range of blood volumes as low as 25 ul, making it suitable for use with venous blood samples or finger-stick blood samples. The assay requires minimal hands-on time for sample preparation, which would facilitate its use in resource-constrained settings.

An earlier model (“NAT2Pred”) predicted NAT2 acetylation phenotype from unphased genomic data; however, it had moderate error rates in distinguishing intermediate from rapid acetylators. 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\textsuperscript{25}. Indeed, striking ethnic differences in the frequencies of SNPs (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\textsuperscript{15, 26, 27}. We trained our SNP classifier with globally representative data, which resulted in the selection and inclusion of the G191A SNP in our model and assay. This is particularly important as Sub-Saharan Africa bears a substantial burden of tuberculosis disease and mortality as well as HIV co-infection, which is independently associated with greater pharmacokinetic variability and tuberculosis treatment toxicity\textsuperscript{28-31}.

The association between acetylation polymorphisms and isoniazid metabolism was first demonstrated in 1959, and their importance was well characterized in subsequent decades through phenotypic descriptions\textsuperscript{32-34}. Subsequent genotypic descriptions confirmed that NAT2 polymorphisms predicted isoniazid early bactericidal activity, and clinical outcomes including hepatotoxicity, relapse and acquisition of drug resistance. Further dosing studies demonstrated that provision of lower doses to slow acetylators and higher doses to rapid acetylators could achieve target concentrations\textsuperscript{35}. One randomized trial of pharmacogenomic-guided dosing of isoniazid during active tuberculosis treatment found that it significantly reduced toxicities (among slow acetylators) and treatment non-response (among rapid acetylators). Taken together, the evidence for pharmacogenomic guided dosing to achieve consistent drug levels and improve clinical outcomes is strong. Automated, easy-to-use assays could enable pharmacogenomic guided isoniazid dosing in resource constrained settings, where a substantial burden of the world’s tuberculosis occurs.
The findings of this study are subject to several limitations. We tested the assays on 49 individuals with active tuberculosis and 20 healthy individuals with a diverse representation of polymorphisms, but the number of participants with G191A mutations was limited (n=5). A larger validation study involving testing on whole blood, including from finger stick capillary blood, is needed to assess real-world performance of this assay under field conditions. Further studies should also investigate testing on non-invasive samples including saliva or oral swabs, from which DNA is abundant. Second, we focused on NAT2 polymorphisms, as they explain 88% of interindividual pharmacokinetic variability, though polymorphisms in several other genes have been associated with hepatotoxicity. However, these associations have been comparatively modest and somewhat inconsistent\textsuperscript{36-40}. We focused on isoniazid and did not include other important tuberculosis drugs, such as rifampicin. The evidence base for pharmacogenomic markers predicting rifampin pharmacokinetics is less robust, and findings concerning clinical outcomes such as toxicities or treatment response are limited\textsuperscript{41,42}. However, given the importance of this drug class in treatment of active and latent tuberculosis, and emerging evidence supporting greater efficacy of higher doses of rifampin, further investigation of pharmacogenomic markers in rifampicin is needed. Future assays may include polymorphisms influencing rifampicin metabolism to further optimize treatment of tuberculosis.

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 unchanged\textsuperscript{43}. Additionally, isoniazid remains a major component of regimens for treatment of latent tuberculosis, which is recommended by the WHO for young children, HIV-infected individuals and household contacts of tuberculosis cases\textsuperscript{44}. More than half the world’s
population have slow or rapid acetylation phenotypes, which put them at risk for excessive drug levels resulting in drug toxicities or insufficient drug levels putting them at risk of acquired drug resistance or disease relapse. Dose adjustment based on NAT2 acetylation genotyping can achieve consistent, target drug levels and reduce the incidence of poor clinical outcomes. We developed a prototype automated, cartridge-based assay that can reliably predict acetylation phenotype directly from as low as 25 ul of whole blood. By developing this for the GeneXpert platform, which is widely used in low- and middle-income countries for tuberculosis diagnosis, this assay could make personalized tuberculosis treatment dosing available in resource-constrained settings. Further studies are needed to evaluate its accuracy and clinical impact in real-world clinical settings.

METHODS

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.

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\textsuperscript{1,25}. Population information on the combined dataset is provided in Supplementary Table 3.

**NAT2 acetylator phenotype prediction classifier**

Phased genomes from 8,561 individuals were used and haplotypes were labeled based on 7 polymorphic sites in the NAT2 gene (191G>A, 282C>T, 341T>C, 481C>T, 590G>A, 803A>G and 857G>A), following an international consensus nomenclature (43). Individuals with two slow haplotypes were considered slow acetylators; those with two rapid haplotypes were considered rapid acetylators; and those with one slow and one rapid acetylator were considered intermediate acetylators. We then constructed an unphased dataset containing only information on whether each sample was wild type for both alleles, homozygous variant for both alleles, or heterozygous. We then split the dataset into two-thirds for a training set and one third for an out-of-sample test set, using sampling stratified by geographic representation to ensure representativeness in the training and test sets. We trained a random forest model on the training set using the caret package in R\textsuperscript{45} and assessed classification performance on the test set. We began with a 7 SNP model and eliminated SNPs in sequential models according to the lowest variable importance factor.

**Sample collection**

Sputum and plasma samples from 49 newly diagnosed patients with active pulmonary tuberculosis were collected at the Federal University of Grande Dourados, Brazil. All 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

Reference NAT2 SNP genotyping assays

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

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

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

**PCR amplification for Sanger DNA sequencing**

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

**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.

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

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

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 R-based Xpose (version 4.7 and higher)\textsuperscript{49,50}. One-compartment model with a first-order absorption with a lognormal distribution for inter-individual variability (IIV) on different PK parameter(s)
as well as an additive and/or proportional model for the residual error were tested for the base model selection. Mixture models with two or three subpopulations representing different clearance rate were then evaluated. The first-order conditional estimation with interaction method (FOCEI) was applied and the model-building procedure and model selection was based on the log-likelihood criterion (the difference in the minimum OFV between hierarchical models was assumed to be Chi-square distributed with degrees of freedom equal to the difference in the number of parameters between models), goodness-of-fit plots (e.g. relevant residuals against time randomly distributed around zero), and scientific plausibility of the model. Visual predictive check was conducted to evaluate whether the final model with estimated fixed-effect parameters and covariates adequately describe data.

Automated NAT2-PGx Multiplex PCR and melt curve analysis

Asymmetric PCR and melt curve analysis were performed on a GeneXpert IV instrument using GeneXpert Dx 4.8 software (Cepheid, Sunnyvale). Flex cartridge-01 (Cepheid) were used to perform automated DNA extraction from whole blood followed by PCR amplification and melt curve analysis to detect SNPs. PCR and melt conditions were optimized using mastermix prepared in house. The NAT2-PGx assay was performed in an 80ul reaction volume (70ul mastermix and 10ul eluted DNA). Briefly, 100ul of whole blood was mixed with 900ul of lysis buffer (Cepheid) in a 1.5ml Eppendorf tube. The sample was vortexed for 2-3 sec and incubated at room temperature for 2min. The entire 1ml whole blood and lysis buffer mix was loaded into sample preparation chamber of flex cart-01 for automated DNA extraction. 70ul PCR mastermix 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
mastermix included (2ul of 2U Aptataq exo-DNA polymerase, 1X Betaine, 1X PCR additive reagent, 1X Aptataq buffer, 8mM MgCl2, 400nM FP, 900nM reverse primers, 500nM of 191-Cy5.5 and 857-Alexa-405, 430nM 282-Alexa-647 and 590-FAM and 300nM 341-Alexa-537 probes). PCR was initiated at holding stage- 50°C for 2min, initial denaturation at 94°C for 2min, followed by 50 cycles of denaturation at 95°C for 15 sec, annealing at 57°C for 30 sec and extension at 65°C for 60 sec. The melting program included three steps: denaturation at 95°C for 1 min, followed by renaturation at 40°C for 3 min and a continuous reading of fluorescence from 40 to 72°C by heating at increments of 0.05°C/sec. The MMCA curve was analyzed using the GXP version 4.8 software.

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.

**Funding**

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.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.
Data availability

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

Code availability

Code and dataset used to develop NAT2 classifier are provided with the manuscript.
Acknowledgements

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

Author contributions

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

Conflicts of Interest

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


22. McInnes G. et al. Drug Response Pharmacogenetics for 200,000 UK Biobank Participants. doi: https://doi.org/10.1101/2020.08.09.243311


42. Database of arylamine N-acetyltransferases (http://nat.mbg.duth.gr)


Figure 1. Predicted \textit{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) isoniazid clearance rates, according to acetylation phenotype predicted from 5 SNPs, measured in sputum samples from 49 patients receiving treatment for active tuberculosis.

Figure 2. Schemata for the automated \textit{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.

Figure 3. Negative derivative transformed melt curves for the five \textit{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 \textit{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).

Figure 4. Effect of whole blood sample volume on melting temperature for wild type and mutant alleles at 5 positions in \textit{NAT2}: \textit{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.
Table 1. Summary of populations included in genomic analysis and their acetylation genotypes.

<table>
<thead>
<tr>
<th>Region</th>
<th>Number of individuals</th>
<th>Slow (n, %)</th>
<th>Intermediate (n, %)</th>
<th>Rapid (n, %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Americas</td>
<td>1,112</td>
<td>432 (39%)</td>
<td>463 (42%)</td>
<td>217 (20%)</td>
</tr>
<tr>
<td>Central and South Asia</td>
<td>588</td>
<td>355 (60%)</td>
<td>198 (34%)</td>
<td>35 (6%)</td>
</tr>
<tr>
<td>East Asia</td>
<td>2,308</td>
<td>340 (15%)</td>
<td>1049 (45%)</td>
<td>919 (40%)</td>
</tr>
<tr>
<td>Europe</td>
<td>3,458</td>
<td>1966 (57%)</td>
<td>1249 (36%)</td>
<td>243 (7%)</td>
</tr>
<tr>
<td>North Africa</td>
<td>44</td>
<td>30 (68%)</td>
<td>10 (23%)</td>
<td>4 (9%)</td>
</tr>
<tr>
<td>sub-Saharan Africa</td>
<td>1,051</td>
<td>450 (43%)</td>
<td>459 (44%)</td>
<td>142 (14%)</td>
</tr>
<tr>
<td>Total</td>
<td>8,561</td>
<td>3573 (42%)</td>
<td>3428 (40%)</td>
<td>1560 (18%)</td>
</tr>
</tbody>
</table>
Table 2. Out-of-sample prediction accuracy of unphased NAT2 SNP data for acetylation phenotype in random forest models. Models were trained with 5,738 individuals and tested on 2,823 individuals. Sens: sensitivity. Spec: specificity.

<table>
<thead>
<tr>
<th>NAT2 SNP Positions</th>
<th>SNP number</th>
<th>Accuracy</th>
<th>95% CI</th>
<th>Sens. Rapid</th>
<th>Spec. Rapid</th>
<th>Sens. Slow</th>
<th>Spec. Slow</th>
</tr>
</thead>
<tbody>
<tr>
<td>191, 282, 341, 481, 590, 803, 857</td>
<td>7</td>
<td>1.000</td>
<td>(0.999-1.000)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>191, 282, 341, 481, 590, 857</td>
<td>6</td>
<td>1.000</td>
<td>(0.999-1.000)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>191, 282, 341, 590, 857</td>
<td>5</td>
<td>1.000</td>
<td>(0.999-1.000)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>282, 341, 590, 857</td>
<td>4</td>
<td>0.978</td>
<td>(0.972-0.983)</td>
<td>0.996</td>
<td>0.988</td>
<td>0.969</td>
<td>0.999</td>
</tr>
<tr>
<td>341, 590, 847</td>
<td>3</td>
<td>0.976</td>
<td>(0.970-0.982)</td>
<td>1.000</td>
<td>0.986</td>
<td>0.967</td>
<td>1.000</td>
</tr>
<tr>
<td>341, 590</td>
<td>2</td>
<td>0.852</td>
<td>(0.838-0.865)</td>
<td>1.000</td>
<td>0.889</td>
<td>0.832</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Table 3. Melting temperature (Tm) values for five NAT2 polymorphisms derived from DNA-probe hybrid melts using single-plex assays validated on 49 pulmonary TB patients

<table>
<thead>
<tr>
<th>NAT2 SNP position</th>
<th>Mutant Total samples analyzed</th>
<th>Wild type Total samples analyzed</th>
<th>Het Total samples analyzed</th>
<th>Mutant Tm ± SD</th>
<th>Wild type Tm ± SD</th>
<th>ΔTm (WT-MT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>191</td>
<td>0</td>
<td>44</td>
<td>5</td>
<td>66.3 ± 0.28</td>
<td>61.92 ± 0.68</td>
<td>4.38</td>
</tr>
<tr>
<td>282</td>
<td>6</td>
<td>24</td>
<td>19</td>
<td>62.81 ± 0.40</td>
<td>66.85 ± 0.34</td>
<td>4.04</td>
</tr>
<tr>
<td>341</td>
<td>12</td>
<td>23</td>
<td>14</td>
<td>66.30 ± 0.48</td>
<td>68.70 ± 0.38</td>
<td>2.40</td>
</tr>
<tr>
<td>590</td>
<td>0</td>
<td>35</td>
<td>14</td>
<td>68.0 ± 0.28</td>
<td>64.37 ± 0.28</td>
<td>3.63</td>
</tr>
<tr>
<td>857</td>
<td>1</td>
<td>41</td>
<td>7</td>
<td>66.4 ± 0.11</td>
<td>62.72 ± 0.29</td>
<td>3.68</td>
</tr>
</tbody>
</table>

SD: Standard deviation
Supplementary Table 1. NAT2 polymorphism profiles of 49 pulmonary TB patients detected from 5-SNP and 7-SNP genotyping assays. Acetylator phenotypes predicted from 5-SNP, 6-SNP and 7-SNP classifiers and plasma INH and acetyl-INH levels recorded at various time-points.

Supplementary Table 2. (a) NAT2 genotype of whole blood samples (n=20) analyzed on 5-plex NAT2-PGx assay and validated on Sanger sequencing. (b) NAT2 genotypes of ten whole blood samples analyzed using 200ul, 100ul, 50ul and 25ul samples.

Supplementary Table 3. Population information on the dataset used to develop 5-SNP classifier.

Supplementary Table 4. List of primers and probes sequences for multiplex assay with their corresponding fluorophores and quenchers.
1. **Sample collection**
   - Collect 1-2 drops of whole blood (approx. 50ul)

2. **Whole blood pretreatment**
   - Mix 950ul lysis buffer, vortex for 2 sec and incubate at RT for 2min

3. **Flex cart01 preparation**
   - Pipette 1ml pretreated sample and 70ul PCR mastermix into the cartridge

4. **Automated DNA extraction & qPCR**
   - Load the cartridge in the GeneXpert machine and start the run
   - Total run time = 140min

5. **Data acquisition**
   - Acquire melt curve data for NAT2 gene polymorphism detection

6. **INH acetylator phenotype prediction**
   - Upload NAT2 polymorphism data on phenotype prediction algorithm
   - INH acetylator phenotype:
     - Rapid acetylator
     - Intermediate acetylator
     - Slow acetylator