

Polymorphisms of promoter and coding regions of the arylamine *N*-acetyltransferase 2 (*NAT2*) gene in the Indonesian population: proposal for a new nomenclature

Rika Yuliwulandari · Qomariyah Sachrowardi · Nao Nishida · Miwa Takasu · Lillian Batubara · Tri Panjiasih Susmiarsih · Jecti Teguh Rochani · Riyani Wikaningrum · Risa Miyashita · Taku Miyagawa · Abdul Salam Mudzakir Sofro · Katsushi Tokunaga

Received: 25 October 2007 / Accepted: 29 November 2007 / Published online: 27 December 2007
© The Japan Society of Human Genetics and Springer 2007

Abstract Polymorphisms of arylamine *N*-acetyltransferase 2 (*NAT2*) are reportedly associated with the risk of drug toxicities and development of various diseases. The present study examined *NAT2* polymorphisms in both promoter and coding regions in the Indonesian population using PCR direct sequencing. The promoter and coding regions of *NAT2* displayed 23 polymorphisms/variations, including eight new ones. Seven haplotypes in the promoter region and six haplotypes in the coding region were inferred. The haplotypes in promoter and coding regions showed limited combinations, and 13 combined haplotypes were inferred. The most frequent haplotypes were U1 (38.9%), U2 (33.5%) in the promoter region and *NAT2**4 (37.3%), *NAT2**6A (36.8%) in the coding region. When converted to predicted phenotypes, the studied population comprised 65.4% rapid acetylators and 35.6% slow acetylators according to bimodal distribution. According to trimodal distribution, frequencies of predicted phenotypes were 13.6, 50.8 and 35.6% for rapid, intermediate and slow acetylators, respectively. Frequencies of *NAT2* alleles for the Indonesian population resembled those of other Southeast Asian populations. We also propose a new *NAT2* nomenclature composed of haplotypes in the promoter region and conventional *NAT2* haplotypes in the coding

region, symbolized by *NAT2**4.U1, *NAT2**4.U2, *NAT2**4.U3, *NAT2**4.U5, *NAT2**4.U6, *NAT2**4.U7, *NAT2**6A.U1, *NAT2**7B.U2, *NAT2**7B.U3, *NAT2**5B.U1, *NAT2**5B.U4, *NAT2**12A.U4 and *NAT2**13.U1.

Keywords *NAT2* · Polymorphisms · Acetylation · Promoter · Indonesian

Abbreviations

| | |
|-------------|-------------------------------|
| <i>NAT2</i> | <i>N</i> -acetyltransferase 2 |
| <i>NAT1</i> | <i>N</i> -acetyltransferase 1 |
| PCR | polymerase chain reaction |
| DME | drug-metabolizing enzyme |
| RA | rapid acetylator |
| IA | intermediate acetylator |
| SA | slow acetylator |
| nt | nucleotide |

Introduction

Polymorphisms of drug-metabolizing enzymes (DMEs) significantly affect the outcomes of treatment and are reportedly associated with risks of developing various diseases such as cancer and neurodegenerative disease (Bandman et al. 2000; Hein 2006; Lilla et al. 2006; Wikman et al. 2001). An early example of polymorphism in a phase II DME is the enzyme arylamine *N*-acetyltransferase 2, which metabolizes several drugs, such as isoniazid, sulfamethazine, procainamide and dapsone, in addition to chemicals and carcinogens (Butcher et al. 2002; Gross et al. 1999).

This enzyme is encoded by *NAT2*, which is located on the short arm of chromosome 8 (8p21.3–23.1) and contains

R. Yuliwulandari · N. Nishida · M. Takasu · R. Miyashita · T. Miyagawa · K. Tokunaga (✉)
Department of Human Genetics,
Graduate School of Medicine, University of Tokyo,
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
e-mail: tokunaga@m.u-tokyo.ac.jp

R. Yuliwulandari · Q. Sachrowardi · L. Batubara · T. P. Susmiarsih · J. T. Rochani · R. Wikaningrum · A. S. M. Sofro
Faculty of Medicine, Yarsi University, Jakarta, Indonesia

an 870-base pair (bp) intronless protein-coding region (Blum et al. 1990; Straka et al. 2006). Genetic polymorphisms in *NAT2* have been described, and 36 alleles have been registered at <http://www.louisville.edu/medschool/pharmacology/NAT.html>.

NAT2 variants are also associated with drug-induced hepatotoxicity in tuberculosis and rheumatoid arthritis patients (Hiratsuka et al. 2002; Huang et al. 2002; Soejima et al. 2007). Impaired drug metabolism can lead to serious adverse drug reactions, including peripheral neuritis, fever and hepatic toxicity (Huang et al. 2002; Hiratsuka et al. 2002), whereas wild-type or rapid metabolism may need a higher dose for disease management (Kubota et al. 2007). Therefore, determination of acetylator status may be essential to optimize individual drug dosing, predict toxicity and improve clinical outcomes of therapy (Weinshilboum 2003).

N-acetylation activity has been investigated in various populations and has been classified in bimodal (rapid or slow) and trimodal (rapid, intermediate or slow) distributions (Hein et al. 2000; Kilbane et al. 1990; Kinzig-Schippers et al. 2005; Parkin et al. 1997). Genotype frequencies of *NAT2* vary considerably among different populations (Evans et al. 2001a). The most frequent *NAT2* allele in Asians is *NAT2**4, compared to *NAT2**5 in Caucasian and African populations (Bell et al. 1993; Lin et al. 1993, 1994). Frequency of the slow acetylator phenotype is low in Asian populations (5–30%), but high in African and Caucasian populations (40–90%) (Butcher et al. 2002; Lee et al. 2002; Woolhouse et al. 1997).

The molecular mechanisms underlying the slow phenotype have been reviewed (Butcher et al. 2002). Studies in bacterial and yeast expression systems have shown that generally, nucleotide substitution in the coding region of *NAT2* results in low activity, decreased expression and unstable proteins with the slow acetylator allele (Blum et al. 1991; Hein et al. 1994; Leff et al. 1999).

Genotype/phenotype relationships in *NAT2* have been extensively studied. However, the relationship between *NAT2* genotype and phenotype is not really understood. Even though most studies showed a high concordance between genotypes and phenotypes, none of them showed 100% concordance. *NAT2**4, *NAT2**12A and *NAT2**13 are related to rapid acetylator phenotypes, whereas *NAT2**5, *NAT2**6 and *NAT2**7 are related to slow acetylator phenotypes (Butcher et al. 2002; Cascorbi et al. 1995; Hickman and Sim 1991; Rihs et al. 2007; Woolhouse et al. 1997). The highest discordance (86%) between genotypes and phenotypes was reported in the Hmong population (Southeast Asian immigrants from Laos since 1975) in Minnesota (Straka et al. 2006). The effects of diet, physiological conditions, bacterial or viral infections affecting metabolic activity of *NAT2* and polymorphisms in the non-coding regulatory regions of

NAT2 were deduced as potential sources of this discordance (Straka et al. 2006).

Indonesia is an archipelago country located in Southeast Asia, with more than 230 million inhabitants according to 2006 estimates. The majority Indonesian ethnicities are Javanese and Sundanese, accounting for 59% of the population.

Most studies focusing on *NAT2* have been based on the genotyping of nucleotide variants in the coding region and ignore the potential effects of the role of promoter polymorphisms in determining *NAT2* phenotypes. In addition, functional elements of promoter region and promoter activity have not been defined. Moreover, there is no report that sufficiently examined *NAT2* polymorphisms in the Indonesian population. The present study focused on the identification of *NAT2* polymorphisms in both promoter and coding regions of *NAT2* in the Indonesian population and analyzed the haplotypes in both regions and the relationships among them.

Materials and methods

Subjects

A group of 212 unrelated healthy Indonesian subjects (96 men, 116 women; mean age, 33.28 years; range, 18–68 years) from the general population was subjected to *NAT2* genotyping. Subjects were interviewed for ethnic background for three generations back and only those of Javanese or Sundanese ethnicity were recruited for the study. Written informed consent in the Indonesian language was obtained from each subject prior to enrollment in this study. All study protocols were approved by the research ethics committees of Yarsi University, Jakarta-Indonesia and Faculty of Medicine, University of Tokyo, Japan.

DNA extraction

Peripheral blood was collected into ethylenediamine tetraacetic acid (EDTA) containing tubes, and DNA was extracted from whole blood samples using the QIAamp DNA blood mini kit (Qiagen, Germany).

NAT2 genotyping

PCR was applied to amplify the promoter and coding regions of *NAT2*. The coding region and 2 kb of the promoter region was determined based on the database searches (<http://genome.ucsc.edu/>) and the previous report (Husain et al. 2007).

Direct sequencing was used for variation screening using 48 samples to detect all *NAT2* polymorphisms/ variations and for genotyping of the rest of the samples. The forward primer X2F and reverse primer X2R were used to amplify the coding region, and forward primers P1F and P2F and reverse primers P1R and P2R were used to amplify promoter regions a and b, respectively (Table 1, Fig. 1). These specific primers were designed using Primer3 free software (<http://frodo.wi.mit.edu/>) and by considering homology of the promoter regions of *NAT1*, *NATP* and *NAT2*. For direct sequencing of the coding region, we used reverse primers X2aR, X2bR, X2cR and X2R for variation screening, and finally only X2bR and X2cR primers were used to genotype all remaining samples, whereas in the promoter region, P1aR, P1bR, P2aR and P2R primers were used (Table 1, Fig. 1).

To amplify the coding region, 25 µl of reaction mixture including 5 ng of genomic DNA, 5 pmol of primers, 2 mM of each dNTP, 5 µl of 10 × PCR buffer containing 2 mM MgCl₂, and 1 u of FastStart *Taq* DNA Polymerase (Roche Applied Science, Mannheim, Germany) were used. Thermal cycling parameters were as follows: initial denaturation for 4 min at 96°C, 40 cycles of 30 s at 96°C, 30 s at 50°C and 2 min at 72°C, ending with final elongation for 5 min at 72°C. Afterwards, PCR product of *NAT2* was verified using electrophoresis on a 2% agarose gel prepared in TAE buffer (40 mM Tris–acetate, 1 mM ethylenediamine tetraacetate; pH 8.0), stained with ethidium bromide and inspected under ultraviolet light. To amplify the promoter region, the same method as described for the coding region was applied with an annealing temperature of 55°C. Direct sequencing was

performed using a commercial kit (BigDye Terminator ver 3.1, Applied Biosystems) with an automated sequencer (ABI 3100). Allele-specific PCR was applied using ASX2F1, ASX2F2 and ASX2R primers to confirm the inferred haplotype in the coding region using the same method as described for coding region amplification.

Statistical analysis

Data were statistically analyzed using the χ^2 to determine whether genotype frequencies are in Hardy–Weinberg equilibrium. Values of $p < 0.05$ were considered statistically significant. Linkage disequilibrium (LD) between alleles at two different loci and haplotype frequencies were inferred using the Haploview 3.32 free software (<http://www.broad.mit.edu/mpg/haploview/index.php>) and SNP-Alyze 3.1 software (Dynacom, Tokyo, Japan) based on the expectation-maximization (EM) algorithm.

Results

In total, 23 polymorphisms/variations were found in the promoter and coding regions of *NAT2* in the Indonesian population samples. In the promoter region, eight single nucleotide polymorphisms (SNPs) with minor allele frequencies of >10%, seven infrequent single nucleotide substitutions, and two insertion polymorphisms were identified, and these included six polymorphisms/variations and two insertions that were newly described. In contrast, the coding region displayed six polymorphic sites with

Table 1 Primers used for amplification and direct sequencing of *NAT2*

| Primer ID | Primer name | Primer sequence | Position |
|-----------|-------------|-------------------------------|----------|
| a | P1 F | 5'-CTCTCCCACACTCAGTCAAAA-3' | 18291035 |
| b | P1 R | 5'-CTCTCCAGTAACAACCTTTAAC-3' | 18292171 |
| c | P1bF | 5'-CCAGGAGTAGGAGGCAATCA-3' | 18291569 |
| d | P1a R | 5'-CCAGTTCCACTTGTGAAGGAC-3' | 18291612 |
| e | P2 F | 5'-TGTACTGGAACAGTAGGTGC-3' | 18292023 |
| f | P2 R | 5'-CTAGGCTGAATGCAATCCTC-3' | 18293116 |
| g | P2bF | 5'-ATACCCATGCCACAGGCTAG-3' | 18292463 |
| h | P2a R | 5'-GGTAACACACAATGTGTCTG-3' | 18292597 |
| i | X2 F | 5'-CATGTAAAAGGGATTCATGCAGT-3' | 18301617 |
| j | X2 R | 5'-CTTGTGGCAAAGTATGGATGGA-3' | 18303188 |
| k | X2a R | 5'-TAATTCCTGCCGTCAATGGT-3' | 18302149 |
| l | X2b R | 5'-GATGAAGCCCACCAAACAGT-3' | 18302507 |
| m | X2c R | 5'-ACAGGCCATCCTTAAAATGC-3' | 18302897 |
| n | ASX2F1* | 5'-TGTTAGGAGGGTATTTTTAC-3' | 18302056 |
| o | ASX2F2* | 5'-TGTTAGGAGGGTATTTTTAT-3' | 18302056 |
| p | ASX2R* | 5'-TCTAGCATGAATCACTCTGC-3' | 18302931 |

Primer IDs a–h were used to amplify the promoter region of *NAT2*. Primer IDs i–p were used to amplify the coding region of *NAT2*. Position indicates the position of the first nucleotide for each primer on chromosome 8 (UCSC database, assembled March 2006). Primers with asterisks were used to amplify specific alleles in the coding region and were obtained from a previous study (Kukongviriyapan et al. 2003)

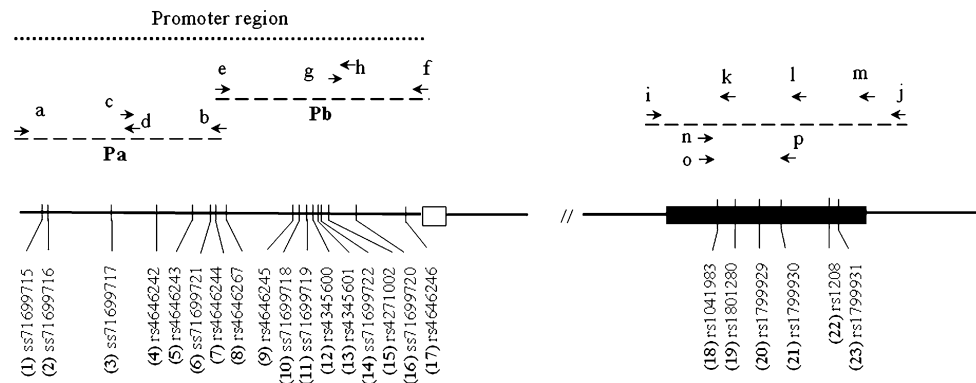


Fig. 1 Gene structure, polymorphic sites and positions of all primers and PCR products in promoter and exon 2 regions of *NAT2*. **a** Open box indicates exon 1 (101 bp), shaded box indicates exon 2 (1,216 bp). **b** Arrows a–p indicate primer positions. **c** Promoter

region (2 kb) was determined from transcription start site according to Husain et al. 2007. **d** Dash lines show PCR products. Pa and Pb represent PCR products in promoter regions a and b, respectively

Table 2 Allele frequencies of observed polymorphisms/variations in promoter and exon 2 regions of *NAT2*

| ID in the present study ^a | NCBI dbSNP ^b | Location | Position on chromosome 8 ^c | Nucleotide changes ^d | MAF (%) | Amino acid changes ^e | pHWE |
|--------------------------------------|-------------------------|----------|---------------------------------------|---------------------------------|---------|---------------------------------|------|
| 1 | ss71699715* | Promoter | 18291165 | C > T | 0.2 | – | NS |
| 2 | ss71699716* | | 18291168 | C > T | 3.1 | – | NS |
| 3 | ss71699717* | | 18291498 | T > C | 0.2 | – | NS |
| 4 | rs4646242 | | 18291729 | A > G | 38.9 | – | NS |
| 5 | rs4646243 | | 18291889 | T > C | 33.7 | – | NS |
| 6 | ss71699721** | | 18291990–1 | ≥ AG | 26.7 | – | NS |
| 7 | rs4646244 | | 18291998 | T > A | 38.9 | – | NS |
| 8 | rs4646267 | | 18292193 | A > G | 33.5 | – | NS |
| 9 | rs4646245 | | 18292393 | T > C | 0.7 | – | 0.01 |
| 10 | ss71699718* | | 18292428 | A > G | 0.9 | – | NS |
| 11 | ss71699719* | | 18292455 | A > T | 0.5 | – | NS |
| 12 | rs4345600 | | 18292488 | A > G | 17.0 | – | NS |
| 13 | rs4345601 | | 18292493 | A > T | 17.0 | – | NS |
| 14 | ss71699722** | | 18292514–5 | ≥ AATT | 26.7 | – | NS |
| 15 | rs4271002 | | 18292548 | G > C | 17.0 | – | NS |
| 16 | ss71699720* | | 18292708 | A > G | 3.8 | – | NS |
| 17 | rs4646246 | | 18292941 | A > G | 34.4 | – | NS |
| 18 | rs1041983 | Exon 2 | 18302075 | T > C | 46.7 | – | NS |
| 19 | rs1801280 | | 18302134 | T > C | 9.0 | I114T | NS |
| 20 | rs1799929 | | 18302274 | C > T | 9.0 | – | NS |
| 21 | rs1799930 | | 18302383 | G > A | 37.0 | R197Q | NS |
| 22 | rs1208 | | 18302596 | A > G | 9.9 | K268R | NS |
| 23 | rs1799931 | | 18302650 | G > A | 15.3 | G286E | NS |

NS Not significant, MAF minor allele frequency, pHWE *p* value for Hardy-Weinberg equilibrium test

^a Numbers 1–23 are the IDs used in the present study

^b Polymorphisms/variations with dbSNP submitter SNP (ss) accession numbers and asterisks show new findings from the present study. Double asterisks indicate insertion polymorphisms

^c Position of polymorphisms/variations on chromosome 8 with reference to the UCSC database, assembled March 2006

^d The first nucleotide in the column of nucleotide changes is the major allele

^e I, isoleucine; T, threonine; R, arginine; Q, glutamine; K, lysine; G, glycine; E, glutamic acid

Table 3 Pairwise LD values between the 16 polymorphisms in the promoter and coding regions of *NAT2*. Upward triangle is pairwise $ID'1$ and downward triangle is pairwise r^2 . The value of each pair was highlighted with the degree of shading corresponding to the strength of LD parameters. Numbers 4–23 represent IDs of polymorphisms used in the present study

| r^2 $ID'1$ | 4 | 5 | 6 | 7 | 8 | 12 | 13 | 14 | 15 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|--------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 4 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.97 | 1 | 1 | 1 | 1 | 1 |
| 5 | 0.32 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.98 | 1 | 1 | 1 | 1 | 1 |
| 6 | 0.23 | 0.18 | | 1 | 1 | 1 | 1 | 1 | 1 | 0.89 | 0.11 | 0.96 | 0.96 | 1 | 0.96 | 0.98 |
| 7 | 1 | 0.32 | 0.23 | | 1 | 1 | 1 | 1 | 1 | 1 | 0.97 | 1 | 1 | 1 | 1 | 1 |
| 8 | 0.32 | 0.99 | 0.18 | 0.32 | | 1 | 1 | 1 | 1 | 1 | 0.98 | 1 | 1 | 1 | 1 | 1 |
| 12 | 0.13 | 0.10 | 0.56 | 0.13 | 0.10 | | 1 | 1 | 1 | 0.80 | 0.75 | 1 | 1 | 1 | 1 | 0.98 |
| 13 | 0.13 | 0.10 | 0.56 | 0.13 | 0.10 | 1 | | 1 | 1 | 0.80 | 0.75 | 1 | 1 | 1 | 1 | 0.98 |
| 14 | 0.23 | 0.18 | 1 | 0.23 | 0.18 | 0.56 | 0.56 | | 1 | 0.89 | 0.11 | 0.96 | 0.96 | 1 | 0.96 | 0.98 |
| 15 | 0.13 | 0.10 | 0.56 | 0.13 | 0.10 | 1 | 1 | 0.56 | | 0.80 | 0.75 | 1 | 1 | 1 | 1 | 0.98 |
| 17 | 0.33 | 0.97 | 0.15 | 0.33 | 0.96 | 0.07 | 0.07 | 0.15 | 0.07 | | 0.99 | 1 | 1 | 1 | 1 | 1 |
| 18 | 0.53 | 0.55 | 0.00 | 0.53 | 0.55 | 0.10 | 0.10 | 0.00 | 0.10 | 0.57 | | 1 | 1 | 1 | 1 | 1 |
| 19 | 0.06 | 0.05 | 0.25 | 0.06 | 0.05 | 0.02 | 0.02 | 0.25 | 0.02 | 0.05 | 0.11 | | 1 | 1 | 1 | 1 |
| 20 | 0.06 | 0.05 | 0.25 | 0.06 | 0.05 | 0.02 | 0.02 | 0.25 | 0.02 | 0.05 | 0.11 | 1 | | 1 | 1 | 1 |
| 21 | 0.91 | 0.30 | 0.21 | 0.91 | 0.29 | 0.12 | 0.12 | 0.21 | 0.12 | 0.31 | 0.52 | 0.06 | 0.06 | | 1 | 1 |
| 22 | 0.07 | 0.06 | 0.28 | 0.07 | 0.06 | 0.02 | 0.02 | 0.28 | 0.02 | 0.06 | 0.12 | 0.90 | 0.90 | 0.06 | | 1 |
| 23 | 0.12 | 0.09 | 0.47 | 0.12 | 0.09 | 0.85 | 0.85 | 0.47 | 0.85 | 0.10 | 0.16 | 0.02 | 0.02 | 0.11 | 0.02 | |

$ID'1$ or r^2

- 0.81-1.00
- 0.51-0.80
- 0.00-0.50

minor allele frequencies >9%, all of which had already been registered in the database (Table 2; Fig. 1). Observed genotype frequencies of all polymorphisms/ variations found in this study were consistent with Hardy–Weinberg equilibrium except for the SNP rs4646245, which was slightly deviated because only 1 minor homozygote and 1 heterozygote were found.

Pairwise $ID'1$ and r^2 values and haplotype frequencies were estimated for polymorphisms with minor allele frequencies $\geq 5\%$ (Table 3). Of the 23 polymorphisms found, 16 were selected for analysis. High $ID'1$ values were observed throughout the entire region of the gene, and no clear boundary of LD structure was identified between promoter and exon 2 regions.

Table 4 shows haplotypes in the promoter and coding regions, and 13 different combined haplotypes were identified in this study. In the promoter region, seven haplotypes composed of ten polymorphisms with minor allele frequency $\geq 10\%$ were inferred and encoded by U1, U2, U3, U4, U5, U6 and U7. The most frequent haplotypes are U1, U2, U3 and U4 with frequency 38.9, 33.5, 16 and 9.7%, respectively, whereas in the coding region, six haplotypes composed of six polymorphisms with minor allele frequency $\geq 10\%$ were inferred and encoded by *NAT2*4*, *NAT2*6A*, *NAT2*7B*, *NAT2*5B*, *NAT2*12A* and *NAT2*13* based on conventional *NAT2* nomenclature (<http://louisville.edu/medschool/pharmacology/NAT.html>). When both haplotype groups were analyzed together, 13 combined haplotypes were inferred and accounted for 99.8% of all haplotypes estimated in the studied population. *NAT2*4.U2* and *NAT2*6A.U1* were predominant with frequencies >30%, followed by *NAT2*7B.U3* and

*NAT2*5B.U4* with a frequency >8%. These most frequent combined haplotypes together comprise 93.9% of all haplotypes.

NAT2 genotypes were determined based on both *NAT2* alleles referred to conventional *NAT2* nomenclature. Frequencies of *NAT2* genotypes and predicted phenotypes both based on bimodal and trimodal distributions are summarized in Table 5.

Discussion

This is the first complete report of *NAT2* genotyping in the Indonesian population. In addition, to the best of our knowledge, this is also the first study to report combined haplotypes of the promoter and coding regions of *NAT2*.

A total of eight new polymorphisms/ variations were found in the promoter region and registered in NCBI dbSNP. Using the EM algorithm, 13 different combined haplotypes comprising six different haplotypes (*NAT2*4*, **5B*, **6A*, **7B*, **12A*, **13*) in the coding region and seven different haplotypes (U1, U2, U3, U4, U5, U6, U7) in the promoter region were inferred (Table 4). Obviously, a limited number of allele combinations (haplotypes) were possible in the promoter region, due to LD among promoter polymorphisms. In addition, most polymorphisms in the promoter and coding regions of *NAT2* showed maximal LD with $ID'1 = 1$. Those polymorphisms seem to have rarely been separated by recombination or recurrent mutation during the history of the Indonesian population. Several couples of polymorphisms in the promoter and coding regions also showed complete LD with an r^2 value of 1.

Table 4 Frequencies of promoter haplotypes, coding region haplotypes and combined haplotypes of *NAT2* in the Indonesian population

| Promoter haplotype | | | | | | | | | | ID based on newly identified haplotype | Frequency (%) |
|-------------------------|-------------------------|----|-------------------------------------|----|---------------------------------------|---|---------------|----|----|--|---------------|
| 4 | 5 | 6 | 7 | 8 | 12 | 13 | 14 | 15 | 17 | | |
| G | T | N | A | A | A | A | N | G | A | U1 | 38.9 |
| A | C | N | T | G | A | A | N | G | G | U2 | 33.5 |
| A | T | I | T | A | G | T | I | C | A | U3 | 16.0 |
| A | T | I | T | A | A | A | I | G | A | U4 | 9.7 |
| A | T | N | T | A | A | A | N | G | A | U5 | 0.9 |
| A | T | I | T | A | G | T | I | C | G | U6 | 0.7 |
| A | C | I | T | A | G | T | I | C | G | U7 | 0.2 |
| Total | | | | | | | | | | | 99.9 |
| Coding region haplotype | | | | | | ID based on conventional <i>NAT2</i> nomenclature | Frequency (%) | | | | |
| 18 | 19 | 20 | 21 | 22 | 23 | | | | | | |
| C | T | C | G | A | G | <i>NAT2*4</i> | 36.9 | | | | |
| C | C | T | G | G | G | <i>NAT2*5B</i> | 9.0 | | | | |
| T | T | C | A | A | G | <i>NAT2*6A</i> | 36.8 | | | | |
| T | T | C | G | A | A | <i>NAT2*7B</i> | 14.9 | | | | |
| C | T | C | G | G | G | <i>NAT2*12A</i> | 0.8 | | | | |
| T | T | C | G | A | G | <i>NAT2*13</i> | 1.2 | | | | |
| Total | | | | | | | 100.0 | | | | |
| Promoter haplotype | Coding region haplotype | | Frequency of combined haplotype (%) | | Proposed new <i>NAT2</i> nomenclature | | | | | | |
| U1 | <i>NAT2*4</i> | | 1.2 | | <i>NAT2*4.U1</i> | | | | | | |
| U2 | <i>NAT2*4</i> | | 33.3 | | <i>NAT2*4.U2</i> | | | | | | |
| U3 | <i>NAT2*4</i> | | 0.9 | | <i>NAT2*4.U3</i> | | | | | | |
| U5 | <i>NAT2*4</i> | | 0.9 | | <i>NAT2*4.U5</i> | | | | | | |
| U6 | <i>NAT2*4</i> | | 0.7 | | <i>NAT2*4.U6</i> | | | | | | |
| U7 | <i>NAT2*4</i> | | 0.2 | | <i>NAT2*4.U7</i> | | | | | | |
| U1 | <i>NAT2*6A</i> | | 36.8 | | <i>NAT2*6A.U1</i> | | | | | | |
| U2 | <i>NAT2*7B</i> | | 0.2 | | <i>NAT2*7B.U2</i> | | | | | | |
| U3 | <i>NAT2*7B</i> | | 15.1 | | <i>NAT2*7B.U3</i> | | | | | | |
| U1 | <i>NAT2*5B</i> | | 0.2 | | <i>NAT2*5B.U1</i> | | | | | | |
| U4 | <i>NAT2*5B</i> | | 8.7 | | <i>NAT2*5B.U4</i> | | | | | | |
| U4 | <i>NAT2*12A</i> | | 0.9 | | <i>NAT2*12A.U4</i> | | | | | | |
| U1 | <i>NAT2*13</i> | | 0.7 | | <i>NAT2*13.U1</i> | | | | | | |
| Total | | | 99.8 | | | | | | | | |

N no insertion, *I* insertion

These results provide useful information for further studies to simplify the genotyping process for *NAT2* in the Indonesian population.

Haplotypes in the promoter region showed limited combinations with haplotypes in the coding region. According to this result, we propose a new nomenclature for *NAT2* haplotypes, as described in Table 4. Interestingly, the same haplotype in the promoter region could be associated with different haplotypes in the coding region. On the other hand, different haplotypes in the promoter

region could also be associated with the same haplotype in the coding region. The common haplotype *NAT2*4* was associated with haplotypes U1, U2, U3, U5, U6 and U7 in the promoter region, *NAT2*7B* haplotype with haplotypes U2, U3, and *NAT2*5B* with haplotypes U1 and U4 (Table 4). In contrast, haplotype U1 was associated with *NAT2*4*, *NAT2*6A*, *NAT2*5B* and *NAT2*13*, haplotype U2 with *NAT2*4* and *NAT2*7B*, haplotype U3 with *NAT2*4* and *NAT2*7B*, and haplotype U4 with *NAT2*5B* and *NAT2*12A*. Each of four major haplotypes of the coding

Table 5 Frequencies of NAT2 genotypes and predicted phenotypes in the Indonesian population

| NAT2 genotypes | n | (%) | Predicted phenotypes | | | |
|----------------|-----|-------|-----------------------|-------|----------------------|-------|
| | | | Trimodal distribution | (%) | Bimodal distribution | (%) |
| *4/*4 | 28 | 13.2 | RA | 13.6 | RA | 64.4 |
| *4/*12A | 1 | 0.4 | | | | |
| *4/*6A | 61 | 28.8 | IA | 50.8 | | |
| *4/*7B | 26 | 12.3 | | | | |
| *4/*5B | 14 | 7.1 | | | | |
| *6A/*13 | 1 | 0.4 | | | | |
| *6A/*12A | 1 | 0.4 | | | | |
| *7B/*13 | 2 | 0.9 | | | | |
| *7B/*12A | 2 | 0.9 | | | | |
| *6A/*6A | 27 | 12.7 | SA | 35.6 | SA | 35.6 |
| *6A/*7B | 23 | 10.8 | | | | |
| *6A/*5B | 16 | 7.5 | | | | |
| *7B/*7B | 3 | 1.4 | | | | |
| *7B/*5B | 6 | 2.8 | | | | |
| *5B/*5B | 1 | 0.4 | | | | |
| Total | 212 | 100.0 | | 100.0 | | 100.0 |

region *NAT2*4*, *NAT2*6A*, *NAT2*7B* and *NAT2*5B* was predominantly associated with one promoter region haplotype, U2, U1, U3 and U4, respectively, accounting for 93.9% in total, whereas 5.9% represent other combined haplotypes. Accordingly, one important question to be answered in the future is the possibility that different haplotypes in the promoter region might display different levels of promoter activity, leading to different phenotypes in the same haplotype of the coding region.

Frequencies of *NAT2* alleles, genotypes and predicted phenotypes based on polymorphisms in the coding region are summarized in Table 5. Major alleles found in the Indonesian population were *NAT2*4* (36.9%) as wild-type, *NAT2*6A* (36.8%), *NAT2*7B* (14.9%) and *NAT2*5B* (9.0%), whereas *NAT2* alleles with frequencies <2% were *NAT2*12A* and *NAT2*13*. The last two alleles are also rarely found in other populations. The major *NAT2* genotypes with frequencies >10% in order were *NAT2*4/*6A* (28.8%), *NAT2*4/*4* (13.2%), *NAT2*6A/*6A* (12.7%), *NAT2*4/*7B* (12.3%) and *NAT2*6A/*7B* (10.8%). The remaining genotypes were in the range of <10%.

Predicted phenotypes deduced from genotypes were classified into bimodal or trimodal distributions based on polymorphisms in the coding region. According to trimodal distribution, phenotypes were predicted as slow, rapid or intermediate acetylators if a genotype comprised two slow alleles, two rapid alleles or heterozygous slow and rapid allele, respectively. When bimodal distribution was considered, rapid acetylator (RA) status was predicted if the genotype comprised at least a rapid allele, and slow acetylator (SA) status was predicted if the genotype comprised

two slow alleles. In trimodal distribution, frequencies of predicted rapid, intermediate and slow acetylators were 13.6, 50.8 and 35.6%, respectively, while in bimodal distribution, frequencies of predicted rapid and slow acetylators were 64.4 and 35.6%, respectively (Table 5).

As reported by Hiratsuka et al. (2002) and Huang et al. (2002), slow acetylator individuals are prone to develop more severe hepatotoxicity than rapid acetylators. In contrast, Kubota et al. (2007) suggested that rapid acetylator individuals should get 1.5 times higher the isoniazid dose than currently recommended to get the proper effect. Therefore, since the prevalence of slow acetylator was significant (35.6%) among the Indonesian population, one should thus consider a probability of hepatic injury in tuberculosis patients on isoniazid treatment. On the other hand, among 13.6% rapid acetylator, one should also consider the probability of them getting a lower dose of drug treatment than what they should get.

The distribution of *NAT2* alleles in various human populations is demonstrated in Table 6. The frequency of predicted slow acetylators in our studied population resembles that in other Southeast Asian populations. In contrast, frequencies of slow acetylators in Indonesian and other Southeast Asian populations are higher than those in Northeast Asian populations including Chinese, Japanese and Koreans (Lee et al. 2002; Chen et al. 2006; Shishikura et al. 2000), but lower than those in Caucasians and Africans (Butcher et al. 2002; Loktionov et al. 2002; Patin et al. 2006).

The investigation of *NAT2* polymorphisms and deduced acetylator status in this population may provide useful

Table 6 Distribution of *NAT2* alleles in various human populations

| Populations | SA ^a | NAT2*4 | NAT2*5 | NAT2*6 | NAT2*7 | Other alleles | n ^b |
|---|-----------------|--------|--------|--------|--------|---------------|----------------|
| Indonesian (this study) | 0.36 | 0.37 | 0.09 | 0.37 | 0.15 | 0.02 | 424 |
| Malay (Zhao et al. 1995) | 0.35 | 0.41 | 0.12 | 0.38 | 0.09 | – | 292 |
| Thai (Kukongviriyapan et al. 2003) | 0.36 | 0.38 | 0.04 | 0.33 | 0.20 | 0.05 | 470 |
| Filipino (Lin et al. 1994) | 0.37 | 0.40 | 0.07 | 0.36 | 0.18 | – | 200 |
| Taiwanese (Huang et al. 2007) | 0.26 | 0.53 | 0.05 | 0.24 | 0.18 | – | 598 |
| Chinese (Chen et al. 2006) | 0.17 | 0.64 | 0.03 | 0.21 | 0.12 | – | 240 |
| Korean (Lee et al. 2002) | 0.10 | 0.66 | 0.02 | 0.20 | 0.12 | 0.01 | 2000 |
| Japanese (Shishikura et al. 2000) | 0.11 | 0.65 | 0.01 | 0.25 | 0.06 | 0.03 | 218 |
| Indian (Zhao et al. 1995) | 0.31 | 0.44 | 0.20 | 0.32 | 0.04 | – | 278 |
| Iranian (Bakayev et al. 2004) | 0.33 | 0.43 | 0.33 | 0.19 | 0.06 | – | 176 |
| Omani (Tanira et al. 2003) | 0.52 | 0.18 | 0.42 | 0.26 | 0.04 | 0.08 | 254 |
| UK Caucasian (Loktionov et al. 2002) | 0.66 | 0.20 | 0.52 | 0.25 | 0.02 | 0.01 | 224 |
| American Caucasian (Butcher et al. 2002) | 0.55 | 0.23 | 0.44 | 0.27 | 0.03 | 0.03 | 444 |
| Sub-Saharan African (Patin et al. 2006) | 0.33 | 0.05 | 0.25 | 0.23 | 0.004 | 0.47 | 530 |
| Black South African (Loktionov et al. 2002) | 0.40 | 0.13 | 0.32 | 0.19 | – | 0.36 | 202 |

^a SA, Predicted slow acetylator phenotype

^b n, Total number of alleles investigated

information for tuberculosis treatment policy, to prevent and control the incidence of drug-induced hepatotoxicity and related adverse effects of chemotherapy, particularly among slow acetylators. Conversely, rapid acetylator status might also affect response to antituberculosis treatment. Further study to investigate the most appropriate dose of antituberculosis treatment in this population is thus warranted.

Conclusion

Our data are in line with published data from other Southeast Asian populations. However, we also examined polymorphisms in the promoter region of *NAT2*, which has not previously been reported. In addition, we suggest that *NAT2* phenotype of each promoter haplotype we found should be examined to provide more accurate *NAT2* phenotypes, instead of conventional prediction. The present results also will be helpful for future epidemiological or clinical studies and for understanding the genetic basis of acetylation polymorphisms in Indonesian. The new method to genotype *NAT2* in a more cost-effective way should be facilitated to identify all mutations in the promoter and coding regions of *NAT2*.

Acknowledgments We would like to thank all volunteers and all teams in Cisarua Respiratory Hospital and all Public Health Services at Serang District, West Java Indonesia who participated in this study. This work was supported by a grant-in-aid for Scientific Research on Priority Areas “Comprehensive Genomics” from the Ministry of Education, Culture, Sports, Science and Technology of Japan and by a grant-in-aid from the Ministry of Health, Labor and Welfare of Japan.

References

- Bakayev VV, Mohammadi F, Bahadori M, Sheikholislami M, Javeri A, Masjedi MR, Velayati AA (2004) Arylamine *N*-acetyltransferase 2 slow acetylator polymorphisms in unrelated Iranian individuals. *Eur J Clin Pharmacol* 60(7):467–471
- Bandmann O, Vaughan JR, Holmans P, Marsden CD, Wood NW (2000) Detailed genotyping demonstrates association between the slow acetylator genotype for *N*-acetyltransferase 2 (*NAT2*) and familial Parkinson’s disease. *Mov Disord* 15:30–35
- Bell DA, Taylor JA, Butler MA, Stephens EA, Wiest J, Brubaker LH, Kadlubar FF, Lucier GW (1993) Genotype/phenotype discordance for human arylamine *N*-acetyltransferase 2 (*NAT2*) reveals a new slow-acetylator allele common in African-Americans. *Carcinogenesis* 14:1689–1692
- Blum M, Grant DM, McBride W, Heim M, Meyer UA (1990) Human *N*-acetyltransferase genes: isolation, chromosomal localisation and functional expression. *DNA Cell Biol* 9:193–203
- Blum M, Demierre A, Grant DM, Heim M, Meyer UA (1991) Molecular mechanism of slow acetylation of drugs and carcinogens in humans. *Proc Natl Acad Sci USA* 88:5237–5241
- Butcher NJ, Boukouvala S, Sim E, Minchin RF (2002) Pharmacogenetics of the arylamine *N*-acetyltransferases. *Pharmacogenomics J* 2:30–42
- Cascorbi I, Drakoulis N, Brockmoller J, Maurer A, Sperling K, Roots I (1995) Arylamine *N*-acetyltransferase (*NAT2*) mutations and their allelic linkage in unrelated Caucasian individuals: correlation with phenotypic activity. *Am J Hum Genet* 57:581–592
- Chen B, Zhang WX, Cai WM (2006) The influence of various genotypes on the metabolic activity of *NAT2* in a Chinese population. *Eur J Clin Pharmacol* 62:355–359
- Evans DA, McLeod HL, Pritchard S, Tariq M, Mobarek A (2001a) Interethnic variability in human drug responses. *Drug Metab Dispos* 29:606–610
- Gross M, Kruisselbrink T, Anderson K, Lang N, McGovern P, Delongchamp R, Kadlubar F (1999) Distribution and concordance of *N*-acetyltransferase genotype and phenotype in an American population. *Cancer Epidemiol Biomarkers Prev* 8:683–692

- Hein DW (2006) *N*-acetyltransferase 2 genetic polymorphism: effects of carcinogen and haplotype on urinary bladder cancer risk. *Oncogene* 25:1649–1658
- Hein DW, Ferguson RJ, Doll MA, Rustan TD, Gray K (1994) Molecular genetics of human polymorphic *N*-acetyltransferase: enzymatic analysis of 15 recombinant wild-type, mutant, and chimeric NAT2 allozymes. *Hum Mol Genet* 3:729–734
- Hein DW, Doll MA, Fretland AJ, Leff MA, Webb SJ, Xiao GH, Devanaboyina US, Nangiu NA, Feng Y (2000) Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphism. *Cancer Epidemiol* 9:29–42
- Hickman D, Sim E (1991) *N*-acetyltransferase polymorphism, comparison of phenotype and genotype in humans. *Biochem Pharmacol* 42:1007–1014
- Hiratsuka M, Kishikawa Y, Takekuma Y, Matsuura M, Narahara K, Inoue T, Hamdy SI, Endo N, Goto J, Mizugaki M (2002) Genotyping of the *N*-acetyltransferase 2 polymorphisms in the prediction of adverse drug reactions to isoniazid in Japanese patients. *Drug Metab Pharmacokinet* 17(4):357–362
- Huang YS, Chern HD, Su WJ, Wu JC, Yang SY, Chang FY, Lee DS (2002) Polymorphism of the *N*-acetyltransferase 2 gene as a susceptibility risk factor for antituberculosis drug-induced hepatitis. *Hepatology* 35:883–889
- Huang CC, Chien WP, Wong RH, Cheng YW, Chen MC, Chou MC, Lee H (2007) NAT2 fast acetylator genotype is associated with an increased risk of colorectal cancer in Taiwan. *Dis Colon Rectum* 50:1–9
- Husain A, Zhang X, Doll MA, States JC, Barker DF, Hein DW (2007) Identification of *N*-Acetyltransferase 2 (NAT2) transcription start sites and quantitation of NAT2-specific mRNA in human tissues. *Drug Metab Dispos* 35:721–727
- Kilbane AJ, Silbart LK, Manis M, Beitins IZ, Weber WW (1990) Human *N*-acetylation genotype determination with urinary caffeine metabolites. *Clin Pharmacol Ther* 47:470–477
- Kinzig-Schippers M, Tomalik-Scharte D, Jetter A, Scheidel B, Jakob V, Rodamer M, Cascorbi I, Doroshenko O, Sorgel F, Fuhr U (2005) Should we use *N*-Acetyltransferase type 2 genotyping to personalized isoniazid doses? *Antimicrob Agents Chemother* 49(5):1733–1738
- Kubota R, Ohno M, Hasunuma T, Iijima H, Azuma J (2007) Dose-escalation study of isoniazid in healthy volunteers with the rapid acetylator genotype of arylamine *N*-acetyltransferase 2. *Eur J Clin Pharmacol* 63:927–933
- Kukongviriyapan V, Prawan A, Tassaneyakul W, Aiemsard J, Warasiha B (2003) Arylamine *N*-acetyltransferase-2 genotypes in the Thai populations. *Br J Clin Pharmacol* 55:278–281
- Lee SY, Lee KA, Ki CS, Kwon OJ, Kim HJ, Chung MP, Suh GY, Kim JW (2002) Complete sequencing of a genetic polymorphism in NAT2 in the Korean populations. *Clin Chem* 48(5):775–777
- Leff MA, Fretland AJ, Doll MA, Hein DW (1999) Novel human *N*-acetyltransferase 2 alleles that differ in mechanism for slow acetylator phenotype. *J Biol Chem* 274:34519–34522
- Lilla C, Verla-Tebit E, Risch A, Jager B, Hoffmeister M, Brenner H, Chang-Claude J (2006) Effect of NAT1 and NAT2 genetic polymorphisms on colorectal cancer risk associated with exposure to tobacco smoke and meat consumption. *Cancer Epidemiol Biomarkers Prev* 15:99–107
- Lin HJ, Han CY, Lin BK, Hardy S (1993) Slow acetylator mutations in the human polymorphic *N*-acetyltransferase gene in 786 Asians, Blacks, Hispanics, and Whites: application to metabolic epidemiology. *Am J Hum Genet* 52:827–834
- Lin HJ, Han CY, Lin BK, Hardy S (1994) Ethnic distribution of slow acetylator mutations in the polymorphic *N*-acetyltransferase (NAT2) gene. *Pharmacogenetics* 4:125–134
- Loktionov A, Moore W, Spencer SP, Vorster H, Nell T, O'Neill IK, Bingham SA, Cummings JH (2002) Differences in *N*-acetylation genotypes between Caucasians and Black South Africans: implications for cancer prevention. *Cancer Detect Prev* 26:15–22
- Parkin DP, Vandenplas S, Botha FJ, Vandenplas ML, Seifart HI, vanHelden PD, van der Walt BJ, Donald PR, van Jaarsveld PP (1997) Trimodality of isoniazid elimination: phenotype and genotype in patients with tuberculosis. *Am J Respir Crit Care Med* 155:1717–1722
- Patin E, Harmant C, Kidd KK, Kidd J, Froment A, Mehdi SQ, Sica L, Heyer E, Quintana-Murci L (2006) Sub-Saharan African coding sequence variation and haplotype diversity at the *NAT2* gene. *Hum Mutat* 27(7):720–730
- Rihs HP, John A, Scherenberg M, Seidel A, Brüning T (2007) Concordance between the deduced acetylation status generated by high-speed real-time PCR based NAT2 genotyping of seven single nucleotide polymorphisms and human NAT2 phenotypes determined by a caffeine assay. *Clin Chim Acta* 376:240–243
- Shishikura K, Hohjoh H, Tokunaga K (2000) Novel allele containing a 190C>T nonsynonymous substitution in the *N*-acetyltransferase (*NAT2*) gene. *Hum Mutat* 15(6):581–584
- Soejima M, Sugiura T, Kawaguchi Y, Kawamoto M, Katsumata Y, Takagi K, Nakajima A, Mitamura T, Mimori A, Hara M, Kamatani N (2007) Association of the diplotype configuration at the *N*-acetyltransferase 2 gene with adverse events with cotrimoxazole in Japanese patients with systemic lupus erythematosus. *Arthritis Res Ther* 9:1–7
- Straka RJ, Burkhardt T, Lang NP, Hadsall KZ, Tsai MY (2006) Discordance between *N*-acetyltransferase 2 phenotype and genotype in a populations of Hmong subjects. *J Clin Pharmacol* 46:802–811
- Tanira MOM, Simsek M, Balushi KA, Lawatia KA, Barawani HA, Bayoumi RA (2003) Distribution of arylamine *N*-acetyltransferase 2 (*NAT2*) genotypes among Omanis. *SQU J Sci Res Med Sci* 5(1–2):9–14
- Weinshilboum R (2003) Inheritance and drug response. *N Engl J Med* 348:529–537
- Wikman H, Thiel S, Jager B, Schmezer P, Spiegelhalder B, Edler L, Dienemann H, Kayser K, Schulz V, Drings P, Bartsch H, Risch A (2001) Relevance of *N*-acetyltransferase 1 and 2 (*NAT1*, *NAT2*) genetic polymorphisms in non-small cell lung cancer susceptibility. *Pharmacogenetics* 11:157–168
- Woolhouse NM, Qureshi MM, Bastaki SM, Patel M, Abdulrazzaq Y, Bayoumi RA (1997) Polymorphic *N*-acetyltransferase (*NAT2*) genotyping of Emiratis. *Pharmacogenetics* 7(1):73–82
- Zhao B, Lee EJ, Wong JY, Yeoh PN, Gong NH (1995) Frequency of mutant *CYP1A1*, *NAT2* and *GSTM1* alleles in normal Indians and Malays. *Pharmacogenetics* 5:275–280