

Variation of the *N-Acetyltransferase 2* Gene in a Romanian and a Kyrgyz Population

Sylvia Rabstein,¹ Klaus Unfried,² Ulrich Ranft,² Thomas Illig,³ Melanie Kolz,³ Hans-Peter Rihs,¹ Chinara Mambetova,⁴ Mariana Vlad,⁵ Thomas Brüning,¹ and Beate Pesch¹

¹Berufsgenossenschaftliches Forschungsinstitut für Arbeitsmedizin, Institute of Ruhr University of Bochum, Bochum, Germany; ²Institut für umweltmedizinische Forschung at Heinrich Heine University Düsseldorf, GmbH, Düsseldorf, Germany; ³GSF-National Research Centre for Environment and Health, Institute of Epidemiology, GmbH, Neuherberg, Germany; ⁴Kyrgyz Scientific Centre of Hematology, Bishkek, Kyrgyzstan; and ⁵Institute of Public Health "Iuliu Moldovan", Cluj, Romania

Abstract

As part of a project on environmental disasters in minority populations, this study aimed to evaluate differences in the sequence of *N-acetyltransferase 2* (*NAT2*) as a metabolic susceptibility gene in yet unexplored ethnicities. Eight single nucleotide polymorphisms (SNP) in the *NAT2* coding region and a variant in the 3' flanking region were analyzed in 290 unrelated Kyrgyz and 140 unrelated Romanians by SNP-specific PCR analysis. The variants 341C, 481T, and 803G were less and 857A more prevalent in Kyrgyz ($P < 0.0001$). The variant at site 857 indicates Asian descent. 282C>T and 590G>A showed no significant variation by ethnicity. 364G>A and 411A>T turned out to be monomorphic. Database comparisons of the *NAT2* minor allele

frequencies support that Romanians belong to Caucasians and Kyrgyz are in between Caucasians and East Asians. The distributions of predicted haplotypes differed significantly between the two ethnicities where the Kyrgyz showed a higher genetic diversity. The haplotype without mutations was more common in Kyrgyz (40.1% in Kyrgyz, 29.3% in Romanians). Accordingly, the imputed slow acetylator phenotype was less prevalent in Kyrgyz (35.2% versus 51.4% in Romanians). We found pronounced ethnic differences in *NAT2* genotypes with yet unknown effect on the health risks for environmental or occupational exposures in minority populations. (Cancer Epidemiol Biomarkers Prev 2006;15(1):138–41)

Introduction

The role of susceptibility genes in the development of chronic diseases is an important issue in occupational and environmental medicine. Arylamine *N*-acetyltransferase 2 (*NAT2*) catalyzes the addition of an acetyl group from acetyl-CoA to a terminal nitrogen on substrates (1). *NAT2* is a gene with a 870-bp coding region mapped to chromosome 8p22. The GenBank accession number X14672 is commonly used as reference sequence (URL: <http://www.ncbi.nlm.nih.gov>). As of June 5, 2003, 36 human *NAT2* alleles were reported based on variations at 15 nucleotide positions (URL: <http://www.louisville.edu/medschool/pharmacology/NAT.html>). We further refer to this database as *NAT* database. An additional single nucleotide polymorphism (SNP) in the 3' flanking region was documented in the SNP500Cancer database of Cancer Genome Anatomy Project (ref. 2; URL: <http://snp500cancer.nci.nih.gov>) and in the database of Perlegen Sciences (ref. 3; URL: <http://genome.perlegen.com>). We further refer to these databases as SNP500 database and Perlegen database. First results of an international project on genetic susceptibility to environmental carcinogens revealed a large variation of allele frequencies in control populations by ethnicity (4). Here, we report on the genetic variation of *NAT2* in a European (Romanian) and Central-Asian (Kyrgyz) study population analyzed as a part

of the European Commission-funded project "Investigation of the Risk of Cyanide in Gold Leaching on Health and Environment in Central Asia and Central Europe" (IRCYL; ref. 5).

Materials and Methods

Study Population. Unrelated subjects were selected for the analyses of *NAT2* gene variants from two population surveys of IRCYL. *NAT2* genotypes were obtained from whole blood samples collected with informed consent. The study was approved by the Ethics Committee of the Kyrgyz Medical Academy and by the Romanian Ministry of Health. In total, 140 Romanian and 290 Kyrgyz subjects were included.

Genotyping of *NAT2* Variants. Genomic DNA from frozen blood samples was prepared at the Kyrgyz Scientific Center of Haematology (Bishkek, Kyrgyzstan) and the Institute of Public Health (Cluj-Napoca, Romania) using the QIAamp DNA Blood Maxi Kit (Qiagen, Hilden, Germany) according to the protocol of the manufacturer. DNA was then shipped to Institut für umweltmedizinische Forschung (Düsseldorf, Germany) and tested for its applicability in PCR assays. *NAT2* variants documented as polymorphic in the SNP500 database were selected for genotyping with the MassARRAY system (Sequenom, San Diego, CA; ref. 6). These variant sites comprised two synonymous polymorphisms (282C>T and 481C>T), six nonsynonymous SNPs (341T>C, 364G>A, 411A>T, 590G>A, 803A>G, and 857G>A), and a C>T polymorphism in 3' untranslated region (rs2552). Two variants (364G>A and 411A>T) turned out to be monomorphic in the IRCYL populations. No assay could be established for the rare polymorphism 191G>A. Genotyping call rates were $\geq 96.8\%$.

Statistical Analyses. *NAT2* analyses were calculated using SAS 8.02 (Cary, NC). Deviations from Hardy-Weinberg

Received 4/13/05; revised 7/5/05; accepted 10/10/05.

Grant support: European Commission 5th Framework Programme for the Project "Investigation of the Risk of Cyanide in Gold Leaching on Health and Environment in Central Asia and Central Europe (IRCYL)" (contract no. ICA2-CT-2000-100036) and Bundesministerium für Bildung und Forschung (German National Genome Research Net).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Sylvia Rabstein, Berufsgenossenschaftliches Forschungsinstitut für Arbeitsmedizin, Institute of Ruhr University of Bochum, Bürkle-de-la-Camp-Platz 1, D-44789 Bochum, Germany. Phone: 49-234-302-4533; Fax: 49-234-302-4505. E-mail: rabstein@bfga.de
Copyright © 2006 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-05-0256

equilibrium were examined with exact tests. The minor allele was defined as the less frequent nucleotide at the polymorphic site for which frequencies are shown with 95% confidence limits. To investigate differences in genotype distributions by χ^2 test, we used a Caucasian and a Chinese population from public databases (SNP500 database and Perlegen database). Haplotypes were inferred using PHASE version 2.0.2 based on coding SNPs with positional information (7, 8).

Further information on genotyping and the prediction of haplotypes can be found on the internet (URL: <http://www.bgfa.ruhr-uni-bochum.de/specials/NAT2.php>).

Results

The distribution of the genotypes and the frequencies of the minor alleles of *NAT2* SNPs in unrelated Kyrgyz and Romanian subjects and in a Caucasian and Chinese reference population are shown in Table 1. All polymorphisms in the IRCYL study populations were in Hardy-Weinberg equilibrium. The Romanians were not significantly different from Caucasians, except a higher 3' untranslated region C allele frequency. This variant was not found in the Chinese. The Kyrgyz were closer to the Chinese than to Caucasians although the Kyrgyz allele frequencies were in between the frequencies of the two populations. The Kyrgyz differed significantly from the Romanians in four of seven SNPs ($P < 0.0001$) with ~20% lower frequencies of 341C, 481C, and 803G. A higher fraction of 857A has been observed among the Kyrgyz (12.1% versus 3.2%, $P < 0.0001$). There were no significant differences in the frequencies of 282C>T and 590G>A ($P = 0.10$ and $P = 0.62$, respectively).

Sixteen Romanians (11.4%) and 47 Kyrgyz (16.2%) were homozygous for the reference sequence. A total of 54 study participants (12.6%) of either Romanian or Kyrgyz origin were estimated to carry a SNP combination not prevalent in the other study group. For all cases with more than one heterozygous SNP, many potential alleles could be related. Expected frequencies of predicted haplotypes based on coding SNPs are reported in Table 2. In analogy to the allele nomenclature [prefix asterisk (*)], we used the prefix number (#) to indicate predicted haplotypes. In both populations, the haplotypes #4, #5B, and #6A account for >80% of the predicted haplotypes. Overall, the distribution of haplotypes in the two study populations was different (permutation test, $P = 0.01$). We found ethnic differences in expected haplotype frequencies of >10% for #4, #5B, and #7B. The latter haplotype was mainly found in Kyrgyz subjects with an expected frequency of 11.6% based on 857A.

We defined rapid acetylators for simplicity and without loss of generality as subjects carrying one or two fast alleles (*4, *11,

or *12) based on the best haplotype pair for each individual. Sixty-eight Romanians (48.6%) and 188 Kyrgyz (64.8%) would be classified as rapid acetylators (data not shown). In addition, we classified the phenotype according to literature using allelic linkage information (9). No differences in the deduction of the acetylation phenotype were found between the haplotype pair-based and literature-based expert rating. However, with expert-based deduction, there remains uncertainty for >50% of the individuals because of more than one possible combination of allele pairs.

Discussion

As part of a project on environmental disasters (5), this study aimed to evaluate ethnic differences in genes involved in susceptibility to environmental agents. We described *NAT2* genotypes, predicted haplotypes, and deduced phenotypes in a Central Asian (Kyrgyz) and a European (Romanian) study population. During the last decade, a high degree of variability in *NAT2* alleles has been found at 15 nucleotide positions and documented in the *NAT* database. The SNP500Cancer Project verified 9 nucleotide positions of these 15 and an additional SNP in the 3' flanking region. A first decision has therefore to be made on the selection of sequence variants in studies of genetic susceptibility (10). We selected all *NAT2* sequence variants which were verified as polymorphic by the SNP500Cancer Project for genotyping with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Two variants (364G>A and 411A>T) turned out to be monomorphic in the Kyrgyz and Romanian study groups. No assay could be established for 191G>A, which characterizes African descent. Cascorbi and Roots (11) critically reviewed pitfalls in *NAT2* genotyping and considered 282C>T and 341T>C in non-Africans to be sufficient to predict the acetylation phenotype. This has at least to be updated for 857G>A, which characterizes Asian descent.

Romanians settle in Central Europe and speak a Latin language from the time of colonization of Dacia by Roman ancestors. They were repeatedly invaded by Turkish-Mongolians. Overall, the *NAT2* SNP distribution in Romanians was similar to other Caucasians with high frequencies of 481T and 803G (SNP500 database; ref. 12). Compared with the distribution of *NAT2* SNPs in a group of Germans, the Romanians had a lower frequency of 341T (13). The Asian-specific SNP at site 857 is slightly more prevalent in Romanians which may reflect Mongolian invasions. The Kyrgyz settle in Central Asia, have been nomads, and speak a Turk language. They have Turkish, Scytho-Siberian, and Mongolian influences. The study population lives at Lake Issyk-Kul near China. The distribution of *NAT2* SNPs in the

Table 1. Minor allele frequencies of *NAT2* polymorphisms in Romanians, Kyrgyz, Caucasians, and Chinese

Sequence variant	Minor allele frequency, % (95% confidence interval)				P , χ^2 test		
	Romanians, $n = 140$	Caucasians,* $n = 31$	Kyrgyz, $n = 290$	Chinese, [†] $n = 24$	Romanians vs Caucasians	Kyrgyz vs Chinese	Romanians vs Kyrgyz
282T	37.1 (25.2-50.3)	30.0 (24.7-35.7)	37.6 (33.6-41.7)	N.d. [‡]	0.52	N.d.	0.10
590A (197Q)	33.9 (22.3-47.0)	28.6 (23.4-34.3)	26.6 (23.0-30.4)	33.3 (20.4-48.4)	0.49	0.49	0.62
341C (114T)	43.5 (31.0-56.7)	37.8 (32.2-43.8)	19.3 (16.2-22.8)	6.3 (1.3-17.2)	0.10	<0.0001	<0.0001
481T	43.3 (30.6-56.8)	38.2 (32.5-44.2)	19.1 (16.5-23.1)	6.5 (1.4-17.9)	0.49	0.08	<0.0001
803G (268K)	40.3 (28.1-53.6)	36.8 (31.1-42.7)	19.8 (16.7-23.3)	N.d.	0.86	N.d.	<0.0001
857A (286E)	3.2 (0.4-11.2)	1.4 (0.4-3.6)	12.1 (9.5-15.0)	6.3 (1.3-17.2)	0.07	0.48	<0.0001
3' Untranslated region C	10.0 (3.8-20.5)	2.5 (1.0-5.1)	1.3 (0.6-2.7)	0 (0-0.7)	0.04	0.42	0.32

*Caucasians investigated for National Cancer Institute Cancer Genome Anatomy Project (<http://snp500cancer.nci.nih.gov>).

[†]Chinese investigated for Perlegen Sciences (<http://genome.perlegen.com>).

[‡]N.d., no data available.

Table 2. Expected frequencies for predicted NAT2 haplotypes in the IRCYL study populations

Haplotype						Expected frequency, % (SE) [†]		
282C>T	341T>C	481C>T	590G>A	803A>G	857G>A		Romanians	Kyrgyz
C	T	C	G	A	G	#4*	29.3 (0.11)	40.1 (0.30)
C	C	T	G	A	G	#5A	3.2 (0.04)	<0.1 (0.03)
C	C	T	G	G	G	#5B	34.3 (0.11)	17.4 (0.07)
C	C	C	G	G	G	#5C	2.5 (0.10)	1.9 (0.07)
T	T	C	A	A	G	#6A	28.6 (0.03)	25.7 (0.24)
C	T	C	A	A	G	#6B	<0.1 (<0.01)	0.6 (0.26)
C	T	C	G	A	A	#7A	0 (<0.01)	0.16 (0.24)
T	T	C	G	A	A	#7B	1.4 (<0.01)	11.6 (0.26)
C	T	T	G	A	G	#11A	0.7 (0.11)	1.7 (0.08)
C	T	C	G	G	G	#12A	<0.1 (0.04)	0.51 (0.04)
T	T	C	G	G	A		0 (<0.01)	<0.1 (0.03)
T	T	C	A	A	A		0 (0)	0.26 (0.29)
T	T	T	A	A	G		<0.1 (0.03)	<0.1 (0.03)

NAT2 haplotype nomenclature with prefix number (#) in analogy to NAT2 allele nomenclature with prefix asterisk ().

[†]Standard error.

Kyrgyz study population resembled more that of Chinese than of Caucasians although the allele frequencies were in between these populations (Perlegen database; ref. 14).

Comparing Romanians with Kyrgyz, there was a slightly larger proportion of Kyrgyz than Romanians which were homozygous carriers of the reference sequence at the investigated sites. The reference sequence was found more common in Asians than in Europeans and is even predominant in Amerindians but rare in Africans (12-17). We found a pronounced variation in four NAT2 polymorphisms which might indicate characteristics of population genetics. There was no obvious association of ethnic differences on whether a SNP results in an exchange of an amino acid or not. The mutations 282C>T and 590G>A showed no significant differences. The synonymous mutation 282C>T is common in a variety of ethnicities with low variation (SNP500 database; refs. 12, 13, 16). The nonsynonymous mutation 590G>A occurred frequently in both Caucasians and Asians but is rare in Amerindians (16). The alleles 341C, 481T, and 803G were less prevalent in Kyrgyz than in Romanians but they occur in higher frequencies in Chinese and even more in Amerindians (Perlegen database; ref. 16). Interestingly, the allele frequencies of these three variants were similar in each population: $\geq 40\%$ in Caucasians, $\sim 20\%$ in Kyrgyz, 6% in Chinese, and $<3\%$ in Ngawbe Indians. The allele 857A is rare in Caucasians. Its prevalence was 12% in Kyrgyz but $>20\%$ in Pacific Rim and Amerindians (16, 18).

Regarding genetic diversity, the majority of subjects in the IRCYL populations were carriers of more than one heterozygous NAT2 variant. For this group, alleles can be determined by cloning and consecutive sequencing of the single DNA strands. With SNP-based PCR methods, haplotypes can only be deduced. Thirteen haplotypes were predicted in the Kyrgyz and nine haplotypes in the Romanian study population, indicating a higher genetic diversity of the Kyrgyz. Haplotypes #4, #5B, and 6A comprised 92% of the Romanian NAT2 variants. These and haplotype #7B with the mutation at site 857 explained 95% of the Kyrgyz variants. The haplotype #5B occurred with a higher frequency in Caucasians and with a lower frequency in the Kyrgyz from Central Asia but was rare in East Asians and Amerindians (12-17).

The human NAT2 gene is supposed to be a susceptibility factor in the metabolism of xenobiotics and, in particular, in the development of bladder cancer where slow acetylators have been associated with an excess risk (19). NAT2 genotyping has been growingly employed to predict the phenotype. Slow acetylators vary by geographic region. About 50% of Caucasians are potential slow acetylators but there is a lower

fraction in Asian populations (20). We also found a lower frequency of predicted slow acetylators in Kyrgyz in comparison with Romanians. In Chinese, there is an even lower fraction of slow acetylators (14, 21). Slow acetylators can be underestimated when only a few SNPs are used to predict the phenotype (10). The degree of misclassification in the prediction of the phenotype from the NAT2 genotype has been estimated to be up to 7% (13, 22-24).

In summary, we confirmed a large ethnic variation of NAT2 gene sequence. From the evolutionary perspective, this points to a selectively neutral character of the NAT2 gene (16). The variant at site 857 characterizes Asian descent. The NAT2 genotype distribution in Romanians was comparable to other Caucasians whereas Kyrgyz had minor allele frequencies in between Caucasians and East Asians. Ethnic differences in metabolic susceptibility genes indicate differences in metabolic pathways with yet unknown effect on the metabolic capacity and disease risks.

Acknowledgments

We thank the Romanian and Kyrgyz IRCYL teams for excellent field work, Tina Müller for statistical assistance, and Lucia Jorge-Nebert and Daniel Nebert for very fruitful comments.

References

- Hein DW, Doll MA, Fretland AJ, et al. Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer Epidemiol Biomarkers Prev* 2000;9:29-42.
- Packer BR, Yeager M, Staats B, et al. SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes. *Nucleic Acids Res* 2004;32:D528-32.
- Hinds DA, Stuve LL, Nilsen GB, et al. Whole-genome patterns of common DNA variation in three human populations. *Science* 2005;307:1072-9.
- Garte S, Gaspari L, Alexandrie AK, et al. Metabolic gene polymorphism frequencies in control populations. *Cancer Epidemiol Biomarkers Prev* 2001;10:1239-48.
- Ranft U, Pesch B, Vogt A. Gold Extraction in Central and Eastern Europe (CEE) and the Commonwealth of Independent States (CIS)—health and environmental risks. Luxembourg: International Center for Studies and Research in Biomedicine; 2005.
- Weidinger S, Klopp N, Wagenpfeil S, et al. Association of a STAT 6 haplotype with elevated serum IgE levels in a population based cohort of white adults. *J Med Genet* 2004;41:658-63.
- Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68:978-89.
- Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003;73:1162-9.
- Roche Diagnostics. User manual RealArt[®] NAT2 LC—PCR Kit. 2004.
- Deitz AC, Rothman N, Rebbeck TR, et al. Impact of misclassification in genotype-exposure interaction studies: example of *N*-acetyltransferase 2 (NAT2), smoking, and bladder cancer. *Cancer Epidemiol Biomarkers Prev* 2004;13:1543-6.

11. Cascorbi I, Roots I. Pitfalls in *N*-acetyltransferase 2 genotyping. *Pharmacogenetics* 1999;9:123–7.
12. Agundez JA, Olivera M, Martinez C, Ladero JM, Benitez J. Identification and prevalence study of 17 allelic variants of the human NAT2 gene in a white population. *Pharmacogenetics* 1996;6:423–8.
13. Cascorbi I, Drakoulis N, Brockmoller J, Maurer A, Sperling K, Roots I. Arylamine *N*-acetyltransferase (NAT2) mutations and their allelic linkage in unrelated Caucasian individuals: correlation with phenotypic activity. *Am J Hum Genet* 1995;57:581–92.
14. Xie HG, Xu ZH, Ou-Yang DS, et al. Meta-analysis of phenotype and genotype of NAT2 deficiency in Chinese populations. *Pharmacogenetics* 1997;7:503–14.
15. Delomenie C, Sica L, Grant DM, Krishnamoorthy R, Dupret JM. Genotyping of the polymorphic *N*-acetyltransferase (NAT2*) gene locus in two native African populations. *Pharmacogenetics* 1996;6:177–85.
16. Jorge-Nebert LF, Eichelbaum M, Griese EU, Inaba T, Arias TD. Analysis of six SNPs of NAT2 in Ngawbe and Embera Amerindians of Panama and determination of the Embera acetylation phenotype using caffeine. *Pharmacogenetics* 2002;12:39–48.
17. Sekine A, Saito S, Iida A, et al. Identification of single-nucleotide polymorphisms (SNPs) of human *N*-acetyltransferase genes NAT1, NAT2, AANAT, ARD1 and L1CAM in the Japanese population. *J Hum Genet* 2001;46:314–9.
18. Lin HJ, Han Ch, Lin BK, Hardy S. Ethnic distribution of slow acetylator mutations in the polymorphic *N*-acetyltransferase (NAT2) gene. *Pharmacogenetics* 1994;4:125–34.
19. Vineis P, Marinelli D, Autrup H, et al. Current smoking, occupation, *N*-acetyltransferase-2 and bladder cancer: a pooled analysis of genotype-based studies. *Cancer Epidemiol Biomarkers Prev* 2001;10:1249–52.
20. Marcus PM, Vineis P, Rothman N. NAT2 slow acetylation and bladder cancer risk: a meta-analysis of 22 case-control studies conducted in the general population. *Pharmacogenetics* 2000;10:115–22.
21. Zhao B, Seow A, Lee EJ, Lee HP. Correlation between acetylation phenotype and genotype in Chinese women. *Eur J Clin Pharmacol* 2000; 56:689–92.
22. Gross M, Kruisselbrink T, Anderson K, et al. Distribution and concordance of *N*-acetyltransferase genotype and phenotype in an American population. *Cancer Epidemiol Biomarkers Prev* 1999;8:683–92.
23. O'Neil WM, Drobitch RK, MacArthur RD, et al. Acetylator phenotype and genotype in patients infected with HIV: discordance between methods for phenotype determination and genotype. *Pharmacogenetics* 2000; 10:171–82.
24. Rothman N, Stewart WF, Caporaso NE, Hayes RB. Misclassification of genetic susceptibility biomarkers: implications for case-control studies and cross-population comparisons. *Cancer Epidemiol Biomarkers Prev* 1993;2:299–303.

Variation of the *N-Acetyltransferase 2* Gene in a Romanian and a Kyrgyz Population

Sylvia Rabstein, Klaus Unfried, Ulrich Ranft, et al.

Cancer Epidemiol Biomarkers Prev 2006;15:138-141.

Updated version Access the most recent version of this article at:
<http://cebp.aacrjournals.org/content/15/1/138>

Cited articles This article cites 22 articles, 8 of which you can access for free at:
<http://cebp.aacrjournals.org/content/15/1/138.full#ref-list-1>

Citing articles This article has been cited by 2 HighWire-hosted articles. Access the articles at:
<http://cebp.aacrjournals.org/content/15/1/138.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://cebp.aacrjournals.org/content/15/1/138>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.