[Frontiers in Bioscience, Scholar, 5, 672-684, January 1, 2013]

Clarifying haplotype ambiguity of NAT2 in multi-national cohorts

Silvia Selinski[1], Meinolf Blaszkewicz[1], Jose A. G. Agundez[2], Carmen Martinez[2], Elena Garcia-Martin[3], Jan G. Hengstler[1], Klaus Golka[1]

[1]Leibniz Research Centre for Working Environment and Human Factors (IfADo), Dortmund, Germany, [2]Department of Pharmacology, University of Extremadura, Badajoz, Spain, [3]Department of Biochemistry and Molecular Biology, University of Extremadura, Badajoz, Spain

TABLE OF CONTENTS

1. ABSTRACT

N-Acetyltransferase 2 (NAT2) is the key enzyme in aromatic amine metabolism. *NAT2* genotyping requires a subsequent determination of the haplotype pairs (formerly: alleles) to derive the acetylation status. The chromosomal phase of the single nucleotide polymorphisms (SNPs) is unclear for about 2/3 of the genotypes. We investigated *NAT2* genotypes of 1,234 bladder cancer cases and 2,207 controls from Germany, Hungary, Pakistan and Venezuela plus 696 further German cancer cases. We reconstructed *NAT2* haplotypes using PHASE v2.1.1. We analysed if the variability of the *NAT2* haplotypes affected the haplotype reconstruction. Furthermore, we compared population haplotype frequencies in three Caucasian control cohorts (German, Hungarian, Spanish), in Pakistanis and Venezuelans and the impact on bladder cancer. We conclude that a common haplotype reconstruction is feasible, enhances precision and reliability. Hungarian controls showed the largest intra-ethnic variability whereas the Pakistanis showed a haplotype distribution typical for Caucasians. The main differences could be observed for the slow haplotypes *5B, *6A and *7B. The association of slow *NAT2* genotypes with bladder cancer risk was most prominent in the Venezuelan study group.

2. INTRODUCTION

Human arylamine *N*-acetyltransferase 2 (NAT2) plays a key role in the metabolism of aromatic amines (1-3). Polymorphisms in the *NAT2* gene were at first discovered as modifiers of blood and urine levels of anti-tuberculosis drugs, in particular, of the ratio of free biologically active isoniazid compared to its mainly acetylated metabolites (4, 5).

Since then, numerous studies have analysed the impact of inter-individual variation in the acetylation capacity on drug efficacy and side effects as well as on cancer susceptibility especially focussing on exposure to carcinogenic NAT2 substrates such as aromatic and heterocyclic amines (2, 6-21). The role of NAT2 in cancer development is debated for a number of tumours (for review see 22); with respect to bladder cancer large studies and meta-analyses show an increased risk for slow acetylators, in particular, if they were exposed to aromatic amines or smoking (6, 15, 23, 24).

Inference of the NAT2 phenotype from the genotype is a non-trivial task as reflected by numerous studies dedicated to this issue (13, 14, 25-36). Commonly a

panel of seven single nucleotide polymorphisms (SNPs), in particular G191A (rs1801279), C282T (rs1041983), T341C (rs1801280), C481T (rs1799929), G590A (rs1799930), A803G (rs1208) and G857A (rs1799931), is used to determine slow and rapid NAT2 haplotypes (formerly: alleles) (31, 34, 37-41) according to the consensus nomenclature (33, 42-44, http://www.louisville.edu/medschool/pharmacology/NAT.html).

Besides simple prediction of slow and rapid acetylation capacity the distinct haplotypes of these seven SNPs are discussed with respect to velocity and substrates (23, 25, 29-31, 41, 45-48). The difficulty in phenotype prediction arises with ambiguity of haplotype assignment as the standard PCR- and RFLP-based methods provide no phase information, i.e. which genetic variants occur together on the same chromosome. Several studies aimed to clarify this situation, however with increasing numbers of studies and, thus, genotypes the decision for the most likely haplotype pair, given the unphased genotype information, might be still difficult for less frequent haplotypes (37-39, 49).

In particular, substantial inter- and even intra-ethnic variation of the seven characteristic *NAT2* SNPs is well known (23, 34, 50-52) and assumed to result from population-specific selective pressures possibly associated with upcoming agriculture and the resulting change of diet and exposure to xenobiotics (49, 51, 53-55) so that haplotype inference between populations might be problematic. Sabbagh and Darlu conclude that a violation of the assumed haplotype patterns, e.g. assuming a haplotype distribution typical for Caucasians, may lead to a bias in haplotype designation from unphased genotypes (56). So, haplotypes could not necessarily be inferred from the literature, especially in non-Caucasian populations due to variations in the linkage disequilibrium patterns, besides further *NAT2* polymorphisms only relevant in a particular population, e.g. (49, 55).

These known population differences prompted us to analyse their impact on haplotype reconstruction and haplotype distributions in healthy controls as well as in bladder cancer cases of different Caucasian and non-Caucasian populations.

In this study, we reconstructed haplotype pairs from 4,337 subjects (Germans, Hungarians, Pakistanis, Venezuelans) evaluating the effect of different haplotype patterns and compared the reconstruction with sequenced *NAT2* haplotypes from a Spanish cohort (37). As inter- and intra-ethnic differences are usually considered in terms of single SNPs instead of haplotypes, we compared population frequencies of the haplotype distribution in three European control cohorts (German, Hungarian, Spanish), in Pakistanis and Venezuelans and evaluated potential differences between German, Hungarian, Pakistani and Venezuelan bladder cancer cases and controls.

3. MATERIALS AND METHODS

3.1. Subjects

Genotypes of 4,337 subjects from Germany (n=3408), Hungary (n=333), Pakistan (n=331) and Venezuela (n=265) were included in the haplotype reconstruction. The German study group comprised 1,212 healthy controls, 140 persons with coxarthrosis or gonarthrosis, 579 persons with a connective tissue disease, 781 urinary bladder cancer cases, 309 head and neck cancer cases, 194 renal cell carcinoma cases and 193 colon cancer cases who had visited the Central Unit Clinical Occupational Medicine of our institute (IfADo) for different purposes or who were members of study cohorts in different hospitals. The Hungarian study group consisted of 61 hospital-based controls and 272 urinary bladder cancer cases, the Pakistani study group encompassed 225 population based and hospital-based controls and 106 urinary bladder cancer cases and the Venezuelan study group comprised 75 hospital-based controls and 190 urinary bladder cancer cases. Thus, we obtained 1,688 controls and 2,649 subjects from the different case groups. German patients with cox- or gonarthrosis or connective tissue diseases (n=719) were used as urinary bladder cancer controls as the genotype distribution showed no deviation from the German healthy controls. The Spanish study group consists of 1,312 healthy controls as described by Agundez *et al.* (37). The sample collection was approved by the local Ethics Committee and by the IRB (institutional review board).

3.2. *NAT2* genotyping

Each subject had donated 10 ml EDTA blood. DNA was extracted out of leukocytes using standard methods (QIAamp DNA Blood Maxikit, Hilden, Germany) and stored at +4°C. *NAT2* genotyping was performed using PCR- and RFLP-based standardized methods (41, 57, 58). A total of seven SNPs, which were adequate to genotype Caucasians for *NAT2* (57), were investigated, namely rs1801279 (G191A), rs1041983 (C282T), rs1801280 (T341C), rs1799929 (C481T), rs1799930 (G590A), rs1208 (A803G) and rs1799931 (G857A). Leukocyte DNA was isolated from a sample of human blood. Amplification of two fragments of DNA with 442 and 559 bp (base pairs) was achieved by means of PCR (polymerase chain reaction). The amplicon from the first PCR was cleaved using three different restriction enzymes, and that of the second PCR with four different restriction enzymes. After subsequent gel electrophoresis with the addition of ethidium bromide, the various DNA fragments were detected in UV light. The results were documented by photography, and the alleles were assigned according to an evaluation scheme (57).

3.3. Statistical analysis

The haplotype analysis was performed using the program PHASE v2.1.1 (59-61). The approach underlying PHASE is a Bayesian haplotype reconstruction method using coalescent-based models to improve the accuracy of haplotypes for unrelated individuals (62). We used the default model for recombination rate variation (63) to estimate the individual haplotype pairs, their probability as well as the sample haplotype frequencies and applied the implemented permutation test for differences in haplotype frequencies that takes the uncertainty of the haplotype reconstruction into account. We tested for differences between German, Hungarian, Pakistani and Venezuelan controls as well as for differences between cases and controls separately in each of the bladder cancer case-

control series. We performed five independent runs with 1,000 main iterations, 1,000 burn-in iterations and a thinning interval of 1. For the permutation test we started with 100 iterations. We chose the best run showing the maximum consistency across five runs and repeated the analysis with the same settings but increased the permutations to 10,000 for the test. Slow and rapid *NAT2* genotypes were deduced according to the consensus *NAT2* gene nomenclature assuming *4, *12 and *13 haplotypes as rapid (44). We calculated odds ratios (OR), 95% confidence intervals (95% CI) and P values of the exact Fisher test of differences between bladder cancer cases and controls separately for each study group using the SAS/STAT software, version 9.2 (SAS Institute Inc., Cary, USA).

4. RESULTS

We reconstructed the haplotype pairs to clarify the chromosomal phase and estimated the sample frequencies to compare the different control groups and to evaluate differences in the modulating effect of NAT2 on bladder cancer in different populations.

4.1. Clarifying haplotype ambiguity

First, we clarified the ambiguous genotypes in the maximum data set of 1,688 healthy subjects and 2,649 cancer cases (mainly urothelial bladder cancer) from Germany, Hungary, Pakistan and Venezuela (Table 1). We discovered 38 different diplotypes, for 22 of them the phase was not clear. Considering the rather frequent genotypes with a frequency of >1% in the controls eight of 12 genotypes could not be determined without phase ambiguity (67%), for 1,096 out of 1,593 controls the haplotype pair remained unclear from PCR/RFLP methods (69%; all controls: 1,151 out of 1,688, 68%). This ambiguity could be clarified completely by haplotype reconstruction with a probability of the reconstructed haplotype pairs of P>0.93. In particular, all but one haplotype pair had a probability of P>0.98, most of them were almost sure (P=1.00).

Comparing the haplotype reconstruction to earlier results, we discovered a single deviation to the haplotype reconstruction and the verification of haplotypes by sequencing in case of the unphased genotype G/G, C/C, C/C, C/T, G/G, A/G, G/G (0021010; *5A/*5C or *5B/*5D). We obtained as the most likely haplotype pair *5A/*5C with P=0.992 (n=3) whereas sequencing yielded *5B/*5D in a Spanish cohort in accordance with the former haplotype reconstruction also with PHASE v2.1.1 and similar models and adjustments (*5A/*5C less probable with P=0.09, *5B/*5D more likely with P=0.91) (37). Further deviations from this study could not be determined, in particular, those genotypes for which PHASE yielded other results than sequencing in the Spanish study group were not present in the current data.

4.2. Investigating ethnic differences in controls

As the German cox- and gonarthrosis patients and persons with a connective tissue disease had the same haplotype distribution as the respective controls (P=0.3500) we used them as additional controls for ethnic differences (Table 2, 3) and bladder cancer (Table 4, 5). The haplotype sample frequencies in German, Hungarian, Pakistani and Venezuelan controls showed clear differences (P=0.0001; Table 2). Comparing these results to the Spanish haplotype distribution further discrepancies became apparent. In particular, the most frequent haplotypes *4, *5B, *6A and *7B showed a considerable inter- and intra-ethnic variability (Table 2). The rapid *4 haplotype was most frequent in Hungarians (28%) and Venezuelans (27%) whereas Germans and Spaniards (both 22%) showed the lowest frequencies. The frequency of *4 in Pakistani controls was similar to the one in Spanish and German controls (24%).

The slow *5B haplotype varied most among Europeans (25% in Hungarians, 43% in Spaniards). *5B was also most frequent in the German (41%) and Pakistani (40%) controls as well as in the Venezuelans though in the latter *5B was clearly less frequent (32%). Similarly the slow *6A haplotype exhibited the most differences between the European study groups (Spain 25%, Germany 29%, Hungary 37%) with the lowest frequencies in Venezuelans (22%). Again, the Pakistani haplotype frequency (28%) was similar to the German one. The slow *7B haplotype was most frequent in Venezuelans (9%) whereas its frequency in Pakistanis (4%), Germans (3%) and Spaniards (1%) was rather low. The Hungarian controls showed again notable deviations from the other Europeans (6%). Remarkably, in the Spanish cohort several of the *14 haplotypes that are rare in Caucasians could be verified. The Venezuelan controls exhibited also a relatively high frequency of the *14B haplotype (0.26%).

Discriminating only between rapid and slow haplotypes yielded a similar high percentage of slow haplotypes that was highest among the German controls and lowest in Venezuelans (Germany 77%, Pakistan 76%, Spain 75%, Hungary 71%, Venezuela 68%). Hence slow acetylators are most common in German controls (60%) and less common in Venezuelans (45%, P=0.0129 all control groups, Table 3). The differences between the German, Hungarian, Pakistani and Spanish controls are less relevant (P=0.3550 excluding the Venezuelan controls).

Separate haplotype reconstruction for each study group did not yield different results. In particular, haplotypes sample frequencies varied at most 0.01 percent points, except for the Venezuelan controls that showed deviation of at most 0.2 percent points.

4.3. Ethnic differences in haplotype distribution with respect to bladder cancer

Finally, we investigated the impact of *NAT2* haplotypes on bladder cancer risk in the German, Hungarian, Pakistani and Venezuelan study groups (Table 4, 5). Differences between cases and controls could only be confirmed for the Venezuelan study group (P=0.0006) whereas German, Hungarian and Pakistani bladder cancer cases showed no significant alterations in haplotype frequency (Table 4). In particular, the Venezuelan controls

Table 1. *NAT2* genotyping yields about 2/3 ambiguous results that could be clarified completely by haplotype reconstruction

	Haplotype pair					Frequency		
Observed diplotype[1]	**Haplotype1**	**Haplotype 2**	**Haplotype 1**	**Haplotype 2**	**P value**	**N total**	**Controls**	**Cancer cases**
0000000	GCTCGAG	GCTCGAG	*4	*4	1.0	216	5.21%	4.83%
0011000	GCTCGAG	GCCTGAG	*4	*5A	1.0	65	1.18%	1,70%
0011010	GCTCGAG	GCCTGGG	*4	*5B	0.998	803	18.48%	18,54%
0010010	GCTCGAG	GCCCGGG	*4	*5C	1.0	27	0.77%	0.53%
0100100	GCTCGAG	GTTCAAG	*4	*6A	1.0	550	11.97%	13,14%
0000100	GCTCGAG	GCTCAAG	*4	*6B	1.0	2	0.06%	0.04%
0100001	GCTCGAG	GTTCGAA	*4	*7B	1.0	66	2.25%	1,06%
0000010	GCTCGAG	GCTCGGG	*4	*12A	1.0	9	0.41%	0.08%
0100000	GCTCGAG	GTTCGAG	*4	*13A	1.0	2	0.12%	0.00%
0022000	GCCTGAG	GCCTGAG	*5A	*5A	1.0	3	0.00%	0.11%
0022010	GCCTGAG	GCCTGGG	*5A	*5B	1.0	73	0.95%	2,15%
0021010	GCCTGAG	GCCCGGG	*5A	*5C	0.992	3	0.00%	0.11%
0111100	GCCTGAG	GTTCAAG	*5A	*6A	1.0	64	1.42%	1,51%
0111001	GCCTGAG	GTTCGAA	*5A	*7B	1.0	5	0.06%	0.15%
0022020	GCCTGGG	GCCTGGG	*5B	*5B	1.0	645	14.93%	14,84%
0021020	GCCTGGG	GCCCGGG	*5B	*5C	1.0	62	1.18%	1,59%
0111110	GCCTGGG	GTTCAAG	*5B	*6A	1.0	1,035	24.76%	23,29%
0111120	GCCTGGG	GTTCAGG	*5B	*6C	1.0	2	0.00%	0.08%
0111011	GCCTGGG	GTTCGAA	*5B	*7B	1.0	122	2.78%	2,83%
0011020	GCCTGGG	GCTCGGG	*5B	*12A	0.999	30	0.59%	0.76%
0111010	GCCTGGG	GTTCGAG	*5B	*13A	0.997	7	0.36%	0.04%
1111010	GCCTGGG	ATTCGAG	*5B	*14B	0.995	1	0.06%	0.00%
1022020	GCCTGGG	ACCTGGG	*5B	*14C	1.0	1	0.06%	0.00%
0020020	GCCCGGG	GCCCGGG	*5C	*5C	1.0	8	0.12%	0.23%
0110110	GCCCGGG	GTTCAAG	*5C	*6A	1.0	45	0.65%	1%
0110011	GCCCGGG	GTTCGAA	*5C	*7B	1.0	4	0.12%	0.08%
0010020	GCCCGGG	GCTCGGG	*5C	*12A	1.0	1	0.06%	0.00%
0200200	GTTCAAG	GTTCAAG	*6A	*6A	1.0	376	8.12%	9,02%
0200210	GTTCAAG	GTTCAGG	*6A	*6C	1.0	2	0.00%	0.08%
0200101	GTTCAAG	GTTCGAA	*6A	*7B	1.0	73	2.07%	1,43%
0100110	GTTCAAG	GCTCGGG	*6A	*12A	0.934	13	0.41%	0.23%
0101110	GTTCAAG	GCTTGGG	*6A	*12C	0.983	1	0.00%	0.04%
0200100	GTTCAAG	GTTCGAG	*6A	*13A	1.0	6	0.06%	0.19%
1200100	GTTCAAG	ATTCGAG	*6A	*14B	1.0	2	0.06%	0.04%
0200002	GTTCGAA	GTTCGAA	*7B	*7B	1.0	7	0.41%	0.00%
0100011	GTTCGAA	GCTCGGG	*7B	*12A	0.999	2	0.12%	0.00%
0200001	GTTCGAA	GTTCGAG	*7B	*13A	1.0	3	0.12%	0.04%
2011010	ACTCGAG	ACCTGGG	*14A	*14C	1.0	1	0.06%	0.00%
					Sum	4,337	1,688	2,649

N total: total number of observed diplotypes, 0: homozygous reference, 1: heterozygous, 2: homozygous variant. [1]Observed diplotypes are shown as the number of variant alleles at each locus (G191A, C282T, T341C, C481T, G590A, A803G, G857A). Ambiguous diplotypes with at least two heterozygous loci are highlighted grey. Considering diplotypes with a frequency in controls of >1% for 69% of all control subjects and eight of twelve diplotypes the corresponding haplotype pairs are not clear.

exhibited more rapid *4 (27% controls *vs.* 22% cases), *12A (2.4% *vs.* 0.05%) and *13A (2.5% *vs.* 0%) haplotypes and slow *7B (9% *vs.* 3%) haplotypes whereas the slow *5A (2% *vs.* 4%), *5B (32% *vs.* 46%) and *5C (1.6% *vs.* 2.7%) haplotypes are more prominent in Venezuelan cases (Table 4). Discriminating only between rapid and slow haplotypes yielded an elevated risk for the slow acetylators (68% *vs.* 78% slow haplotypes, P=0.0258, OR=1.68, 95% CI=1.06-2.70; 45% *vs.* 59% slow acetylators, P=0.0493, OR=1.72, see Table 5).

Differences between controls and bladder cancer cases were also present in Hungarians (more frequent in controls: *4, *6A; more frequent in cases: *5B, *5C; P=0.1524) and Pakistanis (more frequent in controls: *5B, *7B; more frequent in cases: *6A; P=0.1591; see Table 4) but these deviations were not significant. The frequency of the slow acetylators was not elevated in Pakistani cases (76% *vs.* 75% slow haplotypes, P=0.8470, OR=0.95, 95% CI=0.64-1.41; slow acetylators P=0.7039) but more slow haplotypes could be found Hungarian cases (71% *vs.* 78% slow haplotypes, P=0.1241, OR=1.42, 95% CI=0.88-2.25; slow acetylators P=0.2114). The largest study group from Germany showed negligible differences with respect to the rapid *4 (22% *vs.* 23%) and the slow *5B (41% *vs.* 39%) and *6A (29% *vs.* 30%) haplotypes (P=0.2254) and with respect to the frequency of slow haplotypes combined that were present in 77% and 76% of the German controls and cases, respectively (P=0.5218, OR=0.96, 95% CI=0.83-1.10; slow acetylators P=0.8478). Remarkably, the Venezuelan bladder cancer cases showed a much higher frequency of the slow *5B haplotype (46%) and lower *6A frequency (23%) than all other study groups (Hungary and Pakistan ca. 33% each, Germany 39% and 33%) while the frequency of the rapid haplotypes was comparable (22-25%).

5. DISCUSSION

NAT2 is known to be highly polymorphic in different populations (23, 34, 50, 52). The phase II enzyme is involved in the activation and inactivation, respectively, of many drugs and xenobiotics (1-3, 18, 20, 23, 25, 35, 45) and thus discussed as susceptibility factor for a number of tumours (see 22 for review). Several studies and meta-analyses identified NAT2 as risk factor for urinary bladder

Table 2. The haplotype frequencies of *NAT2* differ clearly between German, Hungarian, Spanish, Pakistani and Venezuelan controls

		Total		Germany		Hungary		Pakistan		Venezuela		Spain
Haplotype		**E(Freq)**	**S.E.**	**E(Freq)**	**S.E.**	**E(Freq)**	**S.E.**	**E(Freq)**	**S.E.**	**E(Freq)**	**S.E.**	**Freq**
GCTCGAG	*4	22.75%	0.000219	21.97%	0.000247	27.86%	0.001001	24.42%	0.000650	27.10%	0.000844	22.0%
GCCTGAG	*5A	2.26%	0.000201	2.27%	0.000231	2.47%	0.001002	2.24%	0.000650	2.12%	0.000664	1.0%
GCCTGGG	*5B	39.88%	0.000207	41.10%	0.000234	25.40%	0.001033	39.76%	0.000650	32.34%	0.000856	43.0%
GCCCGGG	*5C	1.50%	0.000031	1.48%	0.000035	0.82%	0.000257	1.78%	0.000070	1.58%	0.000005	0.7%
GCCCGAG	*5D	0.00%	0.000029	0.00%	0.000033	n.p.	n.p.	0.00%	0.000070	n.p.	n.p.	0.3%
GTCTGGG	*5G	0.00%	0.000039	0.00%	0.000030	n.p.	n.p.	n.p.	n.p.	0.00%	0.000324	0.0%
GTTCAAG	*6A	28.68%	0.000105	29.15%	0.000119	36.88%	0.000258	27.78%	0.000099	22.36%	0.000456	25.0%
GCTCAAG	*6B	0.04%	0.000000	0.05%	0.000000	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	2.0%
GTTCAGG	*6C	0.03%	0.000102	0.03%	0.000115	n.p.	n.p.	0.00%	0.000099	0.01%	0.000456	0.1%
GTTCGAA	*7B	3.68%	0.000011	3.00%	0.000008	5.74%	0.000257	4.00%	0.000001	9.47%	0.000083	1.0%
GCTCGGG	*12A	0.78%	0.000217	0.71%	0.000247	0.83%	0.001035	0.02%	0.000650	2.38%	0.000789	2.0%
GTTCGAG	*13A	0.29%	0.000071	0.13%	0.000048	0.00%	0.000257	n.p.	n.p.	2.37%	0.000560	0.3%
ACTCGAG	*14A	0.02%	0.000000	0.03%	0.000000	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	0.4%
ATTCGAG	*14B	0.04%	0.000057	0.03%	0.000033	n.p.	n.p.	n.p.	n.p.	0.26%	0.000442	0.0%
ACCTGGG	*14C	0.04%	0.000035	0.05%	0.000000	n.p.	n.p.	n.p.	n.p.	0.01%	0.000442	0.5%
ATTCAAG	*14D	0.00%	0.000027	0.00%	0.000033	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	0.2%

Total: German, Hungarian, Pakistani and Venezuelan controls, E(Freq): Estimated frequency, S.E.: Standard error, Freq: Frequency, n.p.: not present, Spain: Haplotypes were verified by sequencing. 1% of the sample are haplotypes not present in the table (*12C 0.7%; *14G 0.2%; *12B 0.1%; *5E, *14E, *14F, *14I 0.04% each). Frequent haplotypes with a frequency of >5% in at least one study groups are highlighted grey, Differences in haplotype distribution between German, Hungarian, Pakistani and Venezuelan controls: p = 0.0001 (10,000 iterations)

Table 3. The acetylation status frequencies differ clearly between German, Hungarian, Spanish, Pakistani and Venezuelan controls

	Study group				
Acetylation status	**Germany**	**Hungary**	**Pakistan**	**Venezuela**	**Spain**
Slow	59.55%	50.82%	56.00%	45.26%	56.02%
Intermediate	35.27%	40.98%	39.11%	45.79%	37.96%
Rapid	5.18%	8.20%	4.89%	8.95%	6.02%
N	1,931	61	225	190	1,312

Chi-square test including the Venezuelan controls P=0.0129, Chi-square test excluding the Venezuelan controls P=0.3550

cancer (6, 15, 23, 24, 64) especially in smokers (20, 65-67) and persons occupationally exposed to bladder carcinogens as aromatic amines (9-17, 68-71).

Incidences of urinary bladder cancer vary substantially between different populations due to differences in exogenous and endogenous risk factors. Age-standardised incidence rates vary for the present populations from 2.6/100,000 for Venezuela to 11.6/100,000 for Germany (Hungary: 11.5, Pakistan 3.5) (ASR(W) incidences from IARC GLOBOCAN 2008 http://globocan.iarc.fr). Slow NAT2 status, especially in presence of bladder carcinogens, is a well-known risk factor for urinary bladder cancer that is highly variable between and within ethnicities (50). Recent studies report at most small to moderate bladder cancer risks for slow acetylators ranging from 1.01 (New England study (6)) to 1.15 (genome-wide study on Caucasians (24)) and 1.40-1.45 (meta-analysis (23) and Spanish bladder cancer study (64)) if smoking is not taken into account. The present data showed an odds ratio of 1.04 for all study groups combined with an elevated bladder cancer risk for slow acetylators in Venezuelans and Hungarians but not in the German and Pakistani study group. Taking different distributions of gender, age and smoking habits into account would yield more reliable results but a certain variation in bladder cancer risk is likely to remain due to the variation of further exogenous and endogenous risk factors.

5.1. Haplotype assignment

Studies on NAT2 impact are hampered by unclear haplotype assignment as unphased genotypes of seven SNPs result in more than one possible haplotype pair if more than one of these seven SNPs is heterozygous. Theoretically, for *n* heterozygous SNPs there are $2^{(n-1)}$ possible haplotype pairs, though in fact some of them are rare or have never been observed. For instance, the *NAT2* haplotype pair *5B/*6A results in an unphased genotype with five heterozygous SNPs (*5B: T341C, C481T, A803G; *6A: C282T, G590A) leading to 16 possible haplotype pairs. Moreover, allele frequencies and thus haplotype frequencies and linkage disequilibrium patterns vary across populations (50, 72) so identifying the genotypic phase is not a trivial task and has to be handled with care across populations (56). The present results suggest that for populations that do not vary extremely in their haplotype structure a common reconstruction is feasible and check-ups within the sub-datasets from the same population can be done with less precision. In particular, we observe that also infrequent unphased genotypes result in the same reconstructed haplotype pairs even if the different populations are considered separately.

However, sequencing data showed that the same unphased genotype may in fact result in different haplotype pairs. This insecurity is reflected partially in the probability of the reconstructed haplotype pairs, so it is highly recommended to consider and record this measure of confidence in the estimate. In general, for haplotype

Table 4. Estimated sample frequency of *NAT2* haplotypes in bladder caner cases and controls in four study groups from Germany, Hungary, Pakistan and Venezuela

Study group	Haplotype	NAT2 allele designation	Controls		Cases		
			E(freq)	S.E.	E(freq)	S.E.	Frequency in the sample
Germany							
781 cases	GCTCGAG	*4	21.97%	0.0002	23.17%	0.0004	1211
1,931 controls	GCCTGAG	*5A	2.27%	0.0002	2.51%	0.0003	126
P=0.2254	GCCTGGG	*5B	41.10%	0.0002	38.72%	0.0003	2193
	GCCCGGG	*5C	1.48%	0.0000	1.86%	0.0000	86
	GTCTGGG	*5G	0.00%	0.0001	0.00%	0.0000	0
	GTTCAAG	*6A	29.14%	0.0002	30.27%	0.0002	1599
	GCTCAAG	*6B	0.05%	0.0000	0.00%	0.0000	2
	GTTCAGG	*6C	0.03%	0.0001	0.01%	0.0002	1
	GTTCGAA	*7B	3.00%	0.0000	3.01%	0.0000	163
	GCTCGGG	*12A	0.71%	0.0002	0.45%	0.0004	34
	GTTCGAG	*13A	0.14%	0.0001	0.00%	0.0000	5
	ACTCGAG	*14A	0.03%	0.0000	0.00%	0.0000	1
	ATTCGAG	*14B	0.01%	0.0001	0.00%	0.0000	0
	ACCTGGG	*14C	0.05%	0.0000	0.00%	0.0000	2
	ATTCAAG	*14D	0.01%	0.0001	0.00%	0.0000	1
Hungary							
272 cases	GCTCGAG	*4	27.85%	0.0012	20.82%	0.0010	147
61 controls	GCCTGAG	*5A	2.48%	0.0014	1.48%	0.0008	11
P=0.1524	GCCTGGG	*5B	25.38%	0.0015	32.89%	0.0008	210
	GCCCGGG	*5C	0.82%	0.0006	4.59%	0.0005	26
	GCCCGAG	*5D	0.00%	0.0000	0.01%	0.0005	0
	GTTCAAG	*6A	36.88%	0.0008	33.11%	0.0014	226
	GTTCAGG	*6C	0.01%	0.0007	0.62%	0.0010	3
	GTTCGAA	*7B	5.74%	0.0000	4.96%	0.0001	34
	GCTCGGG	*12A	0.83%	0.0013	0.69%	0.0010	5
	GCTTGGG	*12C	0.00%	0.0005	0.00%	0.0002	0
	GTTCGAG	*13A	0.00%	0.0003	0.64%	0.0009	3
	ATTCGAG	*14B	0.00%	0.0000	0.09%	0.0009	1
	ATTCAAG	*14D	0.00%	0.0000	0.09%	0.0009	0
Pakistan							
106 cases	GCTCGAG	*4	24.44%	0.0000	25.47%	0.0003	164
225 controls	GCCTGAG	*5A	2.22%	0.0001	4.25%	0.0003	19
P=0.1591	GCCTGGG	*5B	39.78%	0.0001	33.96%	0.0003	251
	GCCCGGG	*5C	1.78%	0.0001	1.89%	0.0000	12
	GTTCAAG	*6A	27.78%	0.0001	33.02%	0.0000	195
	GTTCGAA	*7B	4.00%	0.0000	1.42%	0.0000	21
Venezuela							
75 cases	GCTCGAG	*4	27.07%	0.0014	21.95%	0.0018	136
190 controls	GCCTGAG	*5A	2.15%	0.0011	4.05%	0.0018	14
P=0.0006	GCCTGGG	*5B	32.18%	0.0018	45.95%	0.0018	192
	GCCCGGG	*5C	1.58%	0.0002	2.67%	0.0000	10
	GTCTGGG	*5G	0.01%	0.0005	0.00%	0.0000	0
	GTTCAAG	*6A	22.37%	0.0003	22.67%	0.0003	119
	GTTCGAA	*7B	9.47%	0.0003	2.67%	0.0002	40
	GCTCGGG	*12A	2.41%	0.0012	0.05%	0.0018	9
	GTTCGAG	*13A	2.50%	0.0014	0.00%	0.0002	9
	ATTCGAG	*14B	0.13%	0.0013	0.00%	0.0000	1
	ACCTGGG	*14C	0.13%	0.0013	0.00%	0.0000	0
	ATCTGGG	*14E	0.00%	0.0003	0.00%	0.0000	0

E(Freq): Estimated frequency, S.E.: Standard error,P value: Permutation test P value testing for differences in haplotype frequencies between cases and controls (10,000 iterations)

Table 5. Elevated bladder cancer risk can only be confirmed for Venezuelan slow acetylators

Acetylation status	Germany		Hungary		Pakistan		Venezuela	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Slow	59.15%	59.55%	59.56%	50.82%	53.77%	56.00%	58.67%	45.26%
Rapid	40.85%	40.45%	40.44%	49.18%	46.23%	44.00%	41.33%	54.74%
N	781	1,931	272	61	106	225	75	190
Chi-square test P	0.8478		0.2114		0.7039		0.0493	
OR	0.98		1.43		0.91		1.72	
95% CI	0.83-1.16		0.82-2.49		0.57-1.45		0.999-2.95	

All combined: P=0.5617; OR=1.04; 95% CI=0.90-1.21, using the method of Mantel-Haenszel

ambiguity clarification data from different sources, i.e. populations, case-control groups, can be pooled together to obtain a higher precision of the estimates due to a larger sample size and more unambiguous information as long as the haplotype distribution differs not extremely between the populations.

Meanwhile, there are several statistical algorithms and programs available enabling a rapid and accurate haplotype prediction (73-76; for review see 77, 78) though PHASE v2.1 (59, 60) seems to yield consistently excellent results (62, 79-82). The use of

haplotype reconstruction algorithms has become a standard solution in phase assignment, especially for *NAT2* (49, 52-56, 83-85). Furthermore, programs to derive the *NAT2* haplotypes from unphased genotypes are publicly available, e.g. (86, 87).

5.2. Haplotype distribution

The present case-control series showed clear differences in the haplotype distribution between the study groups. Remarkably, the variability between the control groups was as or even more prominent as the differences between bladder cancer cases and controls. This is in accordance with the findings of most studies including different ethnicities or populations from the same ethnic group. Garcia-Martin, for instance, found in a meta-analysis of the seven most common *NAT2* SNPs notable intra-ethnic differences for most SNPs not only in African or Asian study populations but also in Caucasians from the same geographic region (50).

In the Hungarian controls a remarkable low frequency of the *5B (25%) and quite high frequency of *6A (3/%) haplotypes was observed (Table 2) that could not be explained by comparison with geographically nearby populations. Mrozikiewicz *et al.* (88) found in a Polish study group haplotype frequencies quite similar to Caucasian study groups though also with a lower frequency of *5B (33%) similar to Romanians (34%) (85) - a trend that was not observed in Russians (37%) (89), Czechs (39%) and Greeks (38%) (51).

Remarkably the study group from Pakistan exhibits a *NAT2* haplotype distribution very similar to Caucasian study groups from Western and Central Europe. In fact, Pakistanis belong to those Central and South Asian populations that are considered as Indo-Europeans. Generally, *NAT2* haplotypes are highly variable in Central Asian populations (51, 53). Sabbagh *et al.* conclude that Central and South Asian populations exhibit a haplotype distribution in between Europeans and East Asians with a higher frequency of *7B and a virtually absence of *5B - a trend that could not be confirmed in the present Pakistani group (49).

The Venezuelan study group deviated most with respect to its haplotype distribution with several haplotypes being quite similar to the Spanish cohort, e.g. the occurrence of the rare *14 haplotypes which is debated to be due to an admixture of African descendants in the Spanish population (50). The occurrence of the rare *14 haplotypes is also common in South American populations due to South-West European, especially Spanish and Portuguese, descendants (49, 72, 90, 91). Jorge-Nebert *et al.* (52) found in a study of Amerindians from Panama besides over 60% *4 haplotypes, over 20% *7B haplotypes that are also more frequent in our Venezuelan study group (Table 2) but not in Amerindian and admixed populations from Brazil (72, 92). This high frequency of *7B haplotypes was also reported by Fuselli *et al.* (90) generally for American populations with a high intra-ethnic variability and by Sabbagh *et al.* (49) for Asian and Central American populations. Furthermore, Jorge-Nebert *et al.* (52) observe a lower frequency of the slow genotypes than typical for Caucasians (>50%) which is more similar to Asians and other Amerindians (<30%) (51, 52, 90). The *14 haplotypes were also observed quite often in an admixed Brazilian population, in particular in Brazilians of Caucasian and African descent (72). Native Latin Americans showed different haplotype distribution patterns resulting in a much higher percentage of rapid acetylators (75%) (90). Teixeira *et al.* observed in their two Brazilian study groups considerable differences in the haplotype distribution especially for the *4 haplotype (92). Their mainly Amerindian/European study group was quite similar to our Venezuelan group besides the more frequent occurrence of the *7B haplotype in our sample. The *6A haplotype - less frequent in the Venezuelan study group - was almost absent in the Amerindians from Panama (52) but showed a frequency similar to Caucasians in the Brazilian study groups of Teixeira *et al.* (92) while Talbot *et al.* (72) reported lower *6B frequencies. Fuselli *et al.* discussed the occurrence of *6 haplotypes in native Americans to be solely due to an European or African admixture (90). Thus, the results of the Venezuelan study group which was recruited from the general population with no special focus on native Americans were in accordance with an admixed South-American population with a considerable part of Spanish ancestry.

5.3. Haplotype implications

Most studies consider the mutation sites alone. This hampers comparisons of haplotype frequencies, linkage disequilibrium patterns and distributions of slow, intermediate and rapid acetylators. As rapid, reliable and easy-to-handle haplotype reconstruction algorithms are meanwhile standard tools which allow to pool data from different populations - if they are not too distinct and represented by a sufficient number of samples - we strongly recommend their use to provide valuable genetic information.

Reconstructing *NAT2* haplotypes circumvents for the main body of the data the insecurity about the acetylation status assignment as only one instead of several haplotype pairs has to be considered. Furthermore, it provides valuable information about the chromosomal phase which may - in addition to the acetylation status as potential risk factor regarded in most epidemiological studies - help to clarify and understand the role of NAT2 in the development of some cancers. In particular, it is well-known from *in vitro* assays that *NAT2* haplotypes differ in stability, enzymatic activity and substrate affinity (1, 45, 48, 93). Hence different impacts of the *NAT2* haplotypes on the individual susceptibility for some cancers and other diseases via altered metabolic effects are discussed (25, 29, 45, 47).

For the present study groups conclusions on the haplotype basis are difficult due to sample size limitations. Only the German study group is large enough so that differences in haplotype distributions, e.g. between smoking and non-smoking cases might yield interesting results. Remarkably, the Venezuelan bladder cancer cases showed a clearly elevated frequency of the *5B haplotype that is debated to be extremely slow (48).

5.4. Recent advances in NAT2 phenotype prediction: Reducing the genotype to one or two SNPs

Recently, a novel *NAT2* tagSNP (rs1495741) has been identified that predicts the seven SNP inferred NAT2 phenotype with high accuracy (24, 64). Garcia-Closas *et al.* compared rs1495741 genotypes to NAT2 activity measured in lysates of cryopreserved human hepatocytes (64). They conclude that the novel *NAT2* tagSNP predicts with high accuracy the NAT2 phenotype and can be used as a sole marker in populations with European background (64). Selinski *et al.* confirmed the general good performance of this tagSNP with respect to 3,177 *NAT2* and tagSNP genotypes of European, Venezuelan and Pakistani study groups with minor ethnic differences between both genotypic markers predicting the acetylation capacity (94). Furthermore they found a higher specificity of the *NAT2* genotype with respect to *in vivo* phenotypes of 344 Germans measured by the caffeine test (95, 96) though the sensitivity of the tagSNP was indeed excellent but the latter mispredicted several rapid acetylators as slow. In addition, they suggested alternatively a 2-SNP combination (C282T, T341C; rs1041983, rs1801280) for simple acetylation velocity prediction that yielded almost the same results as the *NAT2* genotype. This combination was already suggested in 1995 by Cascorbi *et al.* and in 2003 by Agundez showing excellent results also for Chinese and Japanese study groups (29, 41, 97). In the present data set the haplotypes that would be misclassified by the 2-SNP genotype as rapid (*6B: total 0.04%, Spain 2%; *6E: not present; *6F: not present; *7A: not present; *14A: total 0.02%, Spain 0.4%; *14E: Spain 0.04%, *14I: Spain 0.04%; see Table 2) or slow (*13A: total 0.3%, Spain 0.3%) are also quite rare.

Comparing these findings to the 28 study groups from different populations and ethnicities from Sabbagh *et al.* (49) (in total 7,988 haplotypes) yielded 0.7% haplotypes that would be mispredicted as rapid by the 2-SNP genotype and 1.3% of the haplotypes being mispredicted as slow. In particular, the slow and rare *6B was found in some African populations, Spaniards and UK Caucasians, *7A was present in one African and the Korean study group and *14A was observed in the African, Spanish and Nicaraguan study groups whereas the rapid *13A haplotype was quite frequent in the African study groups (about 4-10%) and also found in Europeans and East Asians.

Thus, for minimizing *NAT2* genotyping effort the use of the tagSNP rs1495741 (64) or the 2-SNP genotype (C282T, T341C) (94) as phenotype predictors might be reasonable as long as no potential substrate-specific differences between *NAT2* haplotypes have to be taken into account or ethnic differences and limited studies demand a careful use. A reduction of genotyping effort would especially be helpful in individualised dosing depending on NAT2 activity, for instance in anti-tuberculosis therapy and in larger epidemiological studies, for instance evaluating individual susceptibility to bladder carcinogens (7, 8, 19).

5.4. Summary

In brief, haplotype reconstruction yielded consistent results across different populations and study groups enabling the analysis of haplotype frequencies with high reliability. We confirmed intra-ethnic differences within European populations with the Hungarian study group differing most while the Pakistani study group was quite similar to the German and Spanish controls. The main differences between the study groups could be observed for the slow haplotypes *5B that was lowest in Hungarian controls (25% *vs.* 32-43%), *6A that was highest in Hungarian (37%) and lowest in Venezuelan controls (22%) and *7B that was highest in Venezuelan controls (9%). The rare *14 haplotypes were found in Spaniards and Venezuelans but also in Germans. The supposed effect of slow *NAT2* genotypes on bladder cancer could only be confirmed for the Venezuelan study group with a lower frequency of rapid *4 (22% *vs.* 27%) and a higher frequency of slow *5B (46% *vs.* 32%) haplotypes in the cases. The not significant results for the three other bladder cancer case-control series are in accordance with current studies and meta-analysis that show small effects of NAT2 if smoking habits and occupational exposure to bladder carcinogens are not taken into account.

6. ACKNOWLEDGEMENTS

The authors thank Ms Doris Dannappel, Ms Marion Page, Ms Kirsten Liesenhoff-Henze, for excellent technical support.

7. REFERENCES

1. DW Hein: N-Acetyltransferase SNPs: emerging concepts serve as a paradigm for understanding complexities of personalized medicine. *Expert Opin Drug Metab Toxicol* 5, 353-366 (2009)

2. DW Hein: Molecular genetics and function of NAT1 and NAT2: role in aromatic amine metabolism and carcinogenesis. *Mutat Res* 506–507, 65-77 (2002)

3. DW Hein: N-Acetyltransferase genetics and their role in predisposition to aromatic and heterocyclic amine-induced carcinogenesis. *Toxicol Lett* 112-113, 349-56 (2000)

4. HB Hughes, JB Biehl, AP Jones, LH Schmidt: Metabolism of isoniazid in man as related to the occurrence of peripheral neuritis. *Am Rev Tuberc* 70, 266-273 (1954)

5. RS Mitchell, JC Bell: Clinical implications of isoniazid blood levels in pulmonary tuberculosis. *N Engl J Med* 257, 1066-1070 (1957)

6. LE Moore, DR Baris, JD Figueroa, M Garcia-Closas, MR Karagas, MR Schwenn, AT Johnson, JH Lubin, DW Hein, CL Dagnall, JS Colt, M Kida, MA Jones, AR Schned, SS Cherala, SJ Chanock, KP Cantor, DT Silverman, N Rothman: GSTM1 null and NAT2 slow acetylation genotypes, smoking intensity and bladder cancer risk: results from the New England bladder cancer study and NAT2 meta-analysis. *Carcinogenesis* 32, 182-189 (2011)

7. RG Hall, RD Leff, T Gumbo: Treatment of active pulmonary tuberculosis in adults: current standards and

recent advances. Insights from the Society of Infectious Diseases Pharmacists. *Pharmacotherapy* 29, 1468-1481 (2009)

8. K Fukino, Y Sasaki, S Hirai, T Nakamura, M Hashimoto, F Yamagishi, K Ueno: Effects of N-acetyltransferase 2 (NAT2), CYP2E1 and glutathione-S-transferase (GST) genotypes on the serum concentrations of isoniazid and metabolites in tuberculosis patients. *J Toxicol Sci* 33, 187-195 (2008)

9. W Weistenhöfer, M Blaszkewicz, HM Bolt, K Golka: N-Acetyltransferase-2 and medical history in bladder cancer cases with a suspected occupational disease (BK 1301) in Germany. *J Toxicol Environ Health A* 71, 906-910 (2008)

10. K Golka, T Seidel, H Dietrich, G Roth, C Rötzel, R Thier, F Geller, T Reckwitz, H Schulze: Occupational and non-occupational risk factors in bladder cancer patients in an industrialized area located in former East-Germany. *Aktuelle Urol* 36, 417-422 (2005)

11. K Golka, V Prior, M Blaszkewicz, HM Bolt: The enhanced bladder cancer susceptibility of NAT2 slow acetylators towards aromatic amines: a review considering ethnic differences. *Toxicol Lett* 2002 128, 229-241 (2002)

12. K Golka, W Weistenhöfer, P Jedrusik, F Geller, M Blaszkewicz, HM Bolt: N-Acetyltransferase 2 phenotype in painters with bladder cancer and controls. *Ann Acad Med Singapore* 30, 464-467 (2001)

13. K Golka, T Reckwitz, M Kempkes, I Cascorbi, M Blaskewicz, SE Reich, I Roots, J Soekeland, H Schulze, HM Bolt: N-Acetyltransferase 2 (NAT2) and glutathione S-transferase µ (GSTM1) in bladder-cancer patients in a highly industrialized area. *Int J Occup Environ Health* 3, 105-110 (1997)

14. K Golka, V Prior, M Blaszkewicz, I Cascorbi, W Schöps, G Kierfeld, I Roots, HM Bolt: Occupational history and genetic N-acetyltransferase polymorphism in urothelial cancer patients of Leverkusen, Germany. *Scand J Work Environ Health* 22, 332-338 (1996)

15. M Garcia-Closas, N Malats, D Silverman, M Dosemeci, M Kogevinas, DW Hein, A Tardon, C Serra, A Carrato, R Garcia-Closas, J Lloreta, G Castaño-Vinyals, M Yeager, R Welch, S Chanock, N Chatterjee, S Wacholder, C Samanic, M Torà, F Fernández, FX Real, N Rothman: NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet* 366, 649-659 (2005)

16. RJ Hung, P Boffetta, P Brennan, C Malaveille, A Hautefeuille, F Donato, U Gelatti, M Spaliviero, D Placidi, A Carta, A Scotto di Carlo, S Porru: GST, NAT, SULT1A1, CYP1B1 genetic polymorphisms, interactions with environmental exposures and bladder cancer risk in a high-risk population. *Int J Cancer* 110, 598-604 (2004)

17. QW Ma, GF Lin, JG Chen, CQ Xiang, WC Guo, K Golka, JH Shen: Polymorphism of N-acetyltransferase 2 (NAT2) gene polymorphism in Shanghai population: occupational and non-occupational bladder cancer patient groups. *Biomed Environ Sci* 17, 291-298 (2004)

18. R Thier, T Brüning, PH Roos, HP Rihs, K Golka, Y Ko, HM Bolt: Markers of genetic susceptibility in human environmental hygiene and toxicology: the role of selected CYP, NAT and GST genes. *Int J Hyg Environ Health* 206, 149-171 (2003)

19. YS Huang, HD Chern, WJ Su, JC Wu, SL Lai, SY Yang, FY Chang, SD Lee: Polymorphism of the N-acetyltransferase 2 gene as a susceptibility risk factor for antituberculosis drug-induced hepatitis. *Hepatology* 35, 883-889 (2002)

20. P Vineis, D Marinelli, H Autrup, J Brockmöller, I Cascorbi, AK Daly, K Golka, H Okkels, A Risch, N Rothman, E Sim, E Taioli: Current smoking, occupation, N-acetyltransferase-2 and bladder cancer: a pooled analysis of genotype-based studies. *Cancer Epidemiol Biomarkers Prev* 10, 1249-1252 (2001)

21. JG Hengstler, M Arand, ME Herrero, F Oesch: Polymorphisms of N-acetyltransferases, glutathione S-transferases, microsomal epoxide hydrolase and sulfotransferases: influence on cancer susceptibility. *Recent Results Cancer Res* 154, 47-85 (1998)

22. JA Agundez: Polymorphisms of human N-acetyltransferases and cancer risk. *Curr Drug Metab* 9, 520-531 (2008)

23. DW Hein: N-Acetyltransferase 2 genetic polymorphism: effects of carcinogen and haplotype on urinary bladder cancer risk. *Oncogene* 25, 1649-1658 (2006)

24. N Rothman, M Garcia-Closas, N Chatterjee, N Malats, X Wu, JD Figueroa, FX Real, D Van Den Berg, G Matullo, D Baris, M Thun, LA Kiemeney, P Vineis, I De Vivo, D Albanes, MP Purdue, T Rafnar, MA Hildebrandt, AE Kiltie, O Cussenot, K Golka, R Kumar, JA Taylor, JI Mayordomo, KB Jacobs, M Kogevinas, A Hutchinson, Z Wang, YP Fu, L Prokunina-Olsson, L Burdett, M Yeager, W Wheeler, A Tardón, C Serra, A Carrato, R Garcia-Closas, J Lloreta, A Johnson, M Schwenn, MR Karagas, A Schned, G Andriole Jr, R Grubb III, A Black, EJ Jacobs, WR Diver, SM Gapstur, SJ Weinstein, J Virtamo, VK Cortessis, M Gago-Dominguez, MC Pike, MC Stern, JM Yuan, DJ Hunter, M McGrath, CP Dinney, B Czerniak, M Chen, H Yang, SH Vermeulen, KK Aben, JA Witjes, RR Makkinje, P Sulem, S Besenbacher, K Stefansson, E Riboli, P Brennan, S Panico, C Navarro, NE Allen, HB Bueno-de-Mesquita, D Trichopoulos, N Caporaso, MT Landi, F Canzian, B Ljungberg, A Tjonneland, F Clavel-Chapelon, DT Bishop, MT Teo, MA Knowles, S Guarrera,

S Polidoro, F Ricceri, C Sacerdote, A Allione, G Cancel-Tassin, S Selinski, JG Hengstler, H Dietrich, T Fletcher, P Rudnai, E Gurzau, K Koppova, SC Bolick, A Godfrey, Z Xu, JI Sanz-Velez, MD Garcia-Prats, M Sanchez, G Valdivia, S Porru, S Benhamou, RN Hoover, JF Fraumeni Jr, DT Silverman, SJ Chanock: A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat Genet* 42, 978-984 (2010)

25. JM Walraven, Y Zang, JO Trent, DW Hein: Structure/function evaluations of single nucleotide polymorphisms in human N-acetyltransferase 2. *Curr Drug Metab* 9, 471-486 (2008)

26. DM Grant: Structures of human arylamine N-acetyltransferases. *Curr Drug Metab* 9, 465-70 (2008)

27. DM Grant: Molecular genetics of the N-acetyltransferases. *Pharmacogenetics* 3, 45-50 (1993)

28. DM Grant, M. Blum and U.A. Meyer: Polymorphisms of N-acetyltransferase genes. *Xenobiotica* 22, 1073-1081 (1992)

29. I Cascorbi, I Roots: Pitfalls in N-acetyltransferase 2 genotyping. *Pharmacogenetics* 9, 123-127 (1999)

30. DW Hein, MA Doll, TD Rustan, RJ Ferguson: Metabolic activation of N-hydroxyarylamines and N-hydroxyarylamides by 16 recombinant human NAT2 allozymes: effects of 7 specific NAT2 nucleic acid substitutions. *Cancer Res* 55, 3531-3536 (1995)

31. DW Hein, RJ Ferguson, MA Doll, TD Rustan, K Gray: 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 (1994)

32. DW Hein, MA Doll, TD Rustan, K Gray, Y Feng, RJ Ferguson, DM Grant: Metabolic activation and deactivation of arylamine carcinogens by recombinant human NAT1 and polymorphic NAT2 acetyltransferases. *Carcinogenesis* 14, 1633-1638 (1993)

33. KPVatsis, WW Weber, DA Bell, JM Dupret, DAP Evans, DM Grant, DW Hein, HJ Lin, UA Meyer, MV Relling, E Sim, T Suzuki, Y Yamazoe: Nomenclature for N-acetyltransferases. *Pharmacogenetics* 5, 1-17 (1995)

34. DA Bell, JA Taylor, MA Butler, EA Stephens, J Wiest, LH Brubaker, FF Kadlubar, GW Lucier: Genotype/phenotype discordance for human arylamine N-acetyltransferase (NAT2) reveals a new slow-acetylator allele common in African-Americans. *Carcinogenesis* 14, 1689-1692 (1993)

35. M Blum, A Demierre, DM Grant, M Heim, UA Meyer: Molecular mechanism of slow acetylation of drugs and carcinogens in humans. *Proc Natl Acad Sci U S A* 88, 5237-41 (1991)

36. D Hickman, E Sim: N-Acetyltransferase polymorphism. Comparison of phenotype and genotype in humans. *Biochem Pharmacol* 42, 1007-1014 (1991)

37. JA Agundez, K Golka, C Martinez, S Selinski, M Blaszkewicz, E Garcia-Martin: Unraveling ambiguous NAT2 genotyping data. *Clin Chem* 54, 1390-1394 (2008)

38. K Golka, M Blaszkewicz, M Samimi, HM Bolt, S Selinski: Reconstruction of N-acetyltransferase 2 haplotypes using PHASE. *Arch Toxicol* 82, 265-270 (2008)

39. HM Bolt, S Selinski, D Dannappel, M Blaszkewicz, K Golka: Re-investigation of the concordance of human NAT2 phenotypes and genotypes. *Arch Toxicol* 79, 196-200 (2008)

40. AC Deitz, N Rothman, TR Rebbeck, RB Hayes, WH Chow, W Zheng, DW Hein, M Garcia-Closas: Impact of misclassification in genotype-exposure interaction studies: example of N-acetyltransferase 2 (NAT2), smoking, and bladder cancer. *Cancer Epidemiol Biomarkers Prev* 13, 1543-1546 (2004)

41. I Cascorbi, N Drakoulis, J Brockmöller, A Maurer, K Sperling, I Roots: Arylamine N-acetyltransferase (NAT2) mutations and their allelic linkage in unrelated Caucasian individuals: correlation with phenotypic activity. *Am J Hum Genet* 57, 581-592 (1995)

42. DW Hein, DM Grant, E Sim: Update on consensus arylamine N-acetyltransferase gene nomenclature. *Pharmacogenetics* 10, 291-292 (2000)

43. DW Hein, S Boukouvala, DM Grant, RF Minchin, E Sim: Changes in consensus arylamine N-acetyltransferase gene nomenclature. *Pharmacogenet Genomics* 18, 367-368 (2008)

44. Arylamine N-acetyltransferase Nomenclature Committee. Update November 1, 2010. http://www.louisville.edu/medschool/pharmacology/NAT.html (accessed Feb 2, 2011)

45. Y Zang, MA Doll, S Zhao, JC States, DW Hein: Functional characterization of single-nucleotide polymorphisms and haplotypes of human N-acetyltransferase 2. *Carcinogenesis* 28, 1665-1671 (2007)

46. DW Hein, MA Doll, AJ Fretland, MA Leff, SJ Webb, GH Xiao, US Devanaboyina, NA Nangju, Y Feng: Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer Epidemiol Biomarkers Prev* 9, 29-42 (2000)

47. JA Agundez, JG Menaya, R Tejeda, F Lago, M Chavez, J Benitez: Genetic analysis of the NAT2 and CYP2D6 polymorphisms in white patients with non-insulin-

dependent diabetes mellitus. *Pharmacogenetics* 6, 465-472 (1996)

48. MA Leff, AJ Fretland, MA Doll, DW Hein: Novel human N-acetyltransferase 2 alleles that differ in mechanism for slow acetylator phenotype. *J Biol Chem* 274, 34519-34522 (1999)

49. A Sabbagh, A Langaney, P Darlu, N Gerard, R Krishnamoorthy, ES Poloni: Worldwide distribution of NAT2 diversity: implications for NAT2 evolutionary history. *BMC Genet* 9, 21 (2008)

50. E Garcia-Martin: Interethnic and intra-ethnic variability of NAT2 single nucleotide polymorphisms. *Curr Drug Metab* 9, 487-497 (2008)

51. F Luca, G Bubba, M Basile, R Brdicka, E Michalodimitrakis, O Rickards, G Vershubsky, L Quintana-Murci, AI Kozlov, A Novelletto: Multiple advantageous amino acid variants in the NAT2 gene in human populations. *PLoS One* 3, e3136 (2008)

52. LF Jorge-Nebert, M Eichelbaum, EU Griese, T Inaba, TD Arias: Analysis of six SNPs of NAT2 in Ngawbe and Embera Amerindians of Panama and determination of the Embera acetylation phenotype using caffeine. *Pharmacogenetics* 12, 39-48 (2002)

53. H Magalon, E Patin, F Austerlitz, T Hegay, A Aldashev, L Quintana-Murci, E Heyer: Population genetic diversity of the NAT2 gene supports a role of acetylation in human adaptation to farming in Central Asia. *Eur J Hum Genet* 16, 243-251 (2008)

54. E Patin, LB Barreiro, PC Sabeti, F Austerlitz, F Luca, A Sajantila, DM Behar, O Semino, A Sakuntabhai, N Guiso, B Gicquel, K McElreavey, RM Harding, E Heyer, L Quintana-Murci: Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes. *Am J Hum Genet* 78, 423-436 (2006)

55. E Patin, C Harmant, KK Kidd, J Kidd, A Froment, SQ Mehdi, L Sica, E Heyer, L Quintana-Murci: Sub-Saharan African coding sequence variation and haplotype diversity at the NAT2 gene. *Hum Mutat* 27, 720 (2006)

56. A Sabbagh, P Darlu: Inferring haplotypes at the NAT2 locus: the computational approach. *BMC Genet* 6, 30 (2005)

57. M Blaszkewicz, D Dannappel, R Thier, J Lewalter: N-Acetyltransferase 2 (genotyping). In: Analyses of hazardous substances in biological materials, vol 9. Special issue: Markers of susceptibility. Eds: J Angerer, M Müller, T Weiss et al. Eds. *Wiley-VCH*, Weinheim, Germany (2004)

58. I Cascorbi, J Brockmöller, PM Mrozikiewicz, S Bauer, R Loddenkemper, I Roots: Homozygous rapid arylamine N-acetyltransferase NAT2 genotype as susceptibility factor for lung cancer. *Cancer Res* 56, 3961-3966 (1996)

59. M Stephens, NJ Smith, P Donnelly: A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68, 978-89 (2001)

60. M. Stephens and P. Donnelly: A comparison of Bayesian methods for haplotype reconstruction. *Am J Hum Genet* 73, 1162-1169 (2003)

61. M Stephens, P Scheet: Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet* 76, 449-462 (2005)

62. J Marchini, D Cutler, N Patterson, M Stephens, E Eskin, E Halperin, S Lin, ZS Qin, HM Munro, GR Abecasis, P Donnelly and the International HapMap Consortium: A comparison of phasing algorithms for trios and unrelated individuals. *Am J Hum Genet* 78, 437-450 (2006)

63. N Li, M Stephens: Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165, 2213-2233 (2003)

64. M Garcia-Closas, DW Hein, D Silverman, N Malats, M Yeager, K Jacobs, MA Doll, JD Figueroa, D Baris, M Schwenn, M Kogevinas, A Johnson, N Chatterjee, LE Moore, T Moeller, FX Real, S Chanock, N Rothman: A single nucleotide polymorphism tags variation in the arylamine N-acetyltransferase 2 phenotype in populations of European background. *Pharmacogenet Genomics* 21, 231-236 (2011)

65. P Boffetta: Tobacco smoking and risk of bladder cancer. *Scand J Urol Nephrol Suppl* 42(S218), 45-54 (2008)

66. MP Zeegers, RA Goldbohm, PA Van Den Brandt: A prospective study on active and environmental tobacco smoking and bladder cancer risk. *Cancer Causes Control* 13, 83-90 (2002)

67. MP Zeegers, FE Tan, E Dorant, PA Van Den Brandt: The impact of characteristics of cigarette smoking on urinary tract cancer risk: a meta-analysis of epidemiologic studies. *Cancer* 89, 630-639 (2000)

68. GL Delclos, SP Lerner: Occupational risk factors. *Scandinavian Journal of Urology and Nephrology Suppl* 42(S218), 58-63 (2008)

69. RC Reulen, E Kellen, F Buntinx, M Brinkman, MP Zeegers: A meta-analysis on the association between bladder cancer and occupation. *Scandinavian Journal of Urology and Nephrology Suppl* 42(S218), 64-78 (2008)

70. I Cascorbi, I Roots, J Brockmöller: Association of NAT1 and NAT2 polymorphisms to urinary bladder

cancer: significantly reduced risk in subjects with NAT1*10. *Cancer Res* 61, 5051-5056 (2001)

71. A Risch, DM Wallace, S Bathers, E Sim: Slow N-acetylation genotype is a susceptibility factor in occupational and smoking related bladder cancer. *Hum Mol Genet* 4, 231-236 (1995)

72. J Talbot, LA Magno, CV Santana, SM Sousa, PR Melo, RX Correa, G Di Pietro, F Rios-Santos: Interethnic diversity of NAT2 polymorphisms in Brazilian admixed populations. *BMC Genet* 11, 87 (2010)

73. L Excoffier, G Laval, S Schneider: Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1, 47-50 (2005)

74. P Scheet, M Stephens: A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* 78, 629-644 (2006)

75. T Niu, ZS Qin, X Xu, JS Liu: Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am J Hum Genet* 70, 157-169 (2002)

76. ZS Qin, T Nui, JS Liu: Partition–ligation–expectation–maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am J Hum Genet* 71, 1242-1247 (2002)

77. RM Salem, J Wessel, NJ Schork: A comprehensive literature review of haplotyping software and methods for use with unrelated individuals. *Hum Genomics* 2, 39-66 (2005)

78. T Niu: Algorithms for inferring haplotypes. *Genet Epidemiol* 27, 334-347 (2004)

79. C Coulonges, O Delaneau, M Girard, H Do, R Adkins, J-L Spadoni, J-F Zagury: Computation of haplotypes on SNPs subsets: advantage of the "global method". *BMC Genet* 7, 50 (2006)

80. L Eronen, F Geerts, H Toivonen: HaploRec: effcient and accurate large-scale reconstruction of haplotypes. *BMC Bioinforma*tics 7, 542 (2006)

81. J Zhang, M Vingron, MR Hoehe: Haplotype reconstruction for diploid populations. *Hum Hered* 59, 144-156 (2005)

82. RM Adkins: Comparison of the accuracy of methods of computational haplotype inference using a large empirical dataset. *BMC Genet* 5, 22 (2004)

83. M Jain, S Kumar, P Lal, A Tiwari, UC Ghoshal, B Mittal: Association of genetic polymorphisms of N-acetyltransferase 2 and susceptibility to esophageal cancer in north Indian population. *Cancer Invest* 25, 340-346 (2007)

84. SC Sak, JH Barrett, AB Paul, DT Bishop, AE Kiltie: DNA repair gene XRCC1 polymorphisms and bladder cancer risk. *BMC Genet* 8, 13 (2007)

85. S Rabstein, K Unfried, U Ranft, T Illig, M Kolz, HP Rihs, C Mambetova, M Vlad, T Brüning, B Pesch: Variation of the N-acetyltransferase 2 gene in a Romanian and a Kyrgyz population. *Cancer Epidemiol Biomarkers Prev* 15, 138-141 (2006)

86. IB Kuznetsov, M McDuffie, R Moslehi: A web server for inferring the human N-acetyltransferase-2 (NAT2) enzymatic phenotype from NAT2 genotype. *Bioinformatics* 25, 1185-1186 (2009)

87. A Sabbagh, P Darlu, M Vidaud: Evaluating NAT2PRED for inferring the individual acetylation status from unphased genotype data. *BMC Med Genet* 10, 148 (2009)

88. PM Mrozikiewicz, I Cascorbi, J Brockmöller, I Roots: Determination and allelic allocation of seven nucleotide transitions within the arylamine N-acetyltransferase gene in the Polish population. *Clin Pharmacol Ther* 59, 376-382 (1996)

89. EA Gaikovitch, I Cascorbi, PM Mrozikiewicz, J Brockmöller, R Frötschl, K Köpke, T Gerloff, JN Chernov, I Roots: Polymorphisms of drug-metabolizing enzymes CYP2C9, CYP2C19, CYP2D6, CYP1A1, NAT2 and of P-glycoprotein in a Russian population. *Eur J Clin Pharmacol* 59, 303-312 (2003)

90. S Fuselli, RH Gilman, SJ Chanock, SL Bonatto, G De Stefano, CA Evans, D Labuda, D Luiselli, FM Salzano, G Soto, G Vallejo, A Sajantila, D Pettener, E Tarazona-Santos: Analysis of nucleotide diversity of NAT2 coding region reveals homogeneity across Native American populations and high intra-population diversity. *Pharmacogenomics J* 7, 144-152 (2007)

91. C Martinez, JA Agundez, M Olivera, A Llerena, R Ramirez, M Hernández, J Benitez: Influence of genetic admixture on polymorphisms of drug-metabolizing enzymes: analyses of mutations on NAT2 and C gamma P2E1 genes in a mixed Hispanic population. *Clin Pharmacol Ther* 63, 623-628 (1998)

92. RL Teixeira, AB Miranda, AG Pacheco, MQ Lopes, J Fonseca-Costa, MF Rabahi, HM Melo, AL Kritski, FC Mello, PN Suffys, AR Santos: Genetic profile of the arylamine N-acetyltransferase 2 coding gene among individuals from two different regions of Brazil. *Mutat Res* 624, 31-40 (2007)

93. L Le Marchand, L Sivaraman, AA Franke, LJ Custer, LR Wilkens, AF Lau, RV Cooney: Predictors of N-acetyltransferase activity: should caffeine phenotyping and NAT2 genotyping be used interchangeably in epidemiological studies? *Cancer Epidemiol Biomarkers Prev* 5, 449-455 (1996)

94. S Selinski, M Blaszkewicz, ML Lehmann, D Ovsiannikov, O Moormann, C Guballa, A Kress, MC Truß, H Gerullis, T Otto, D Barski, G Niegisch, P Albers, S

Frees, W Brenner, JW Thüroff, M Angeli-Greaves, T Seidel, G Roth, H Dietrich, R Ebbinghaus, HM Prager, HM Bolt, M Falkenstein, A Zimmermann, T Klein, T Reckwitz, HC Roemer, D Löhlein, W Weistenhöfer, W Schöps, SAH Rizvi, M Aslam, G Banfi, I Romics, M Steffens, AB Ekici, A Winterpacht, K Ickstadt, H Schwender, JG Hengstler, K Golka: Genotyping NAT2 with only two SNPs (rs1041983 and rs1801280) outperforms the tagging SNP rs1495741 and is equivalent to the conventional 7-SNP NAT2 genotype. *Pharmacogenet Genomics* 21, 673-678 (2011)

95. DM Grant, BK Tang, W Kalow: A simple test for acetylator phenotype using caffeine. *Br J Clin Pharmacol* 17, 459-464 (1984)

96. M Blaszkewicz: N-Acetyltransferase 2 (phenotyping: caffeine test) In: Analyses of hazardous substances in biological materials, vol 9. Special issue: Markers of susceptibility. Eds: J Angerer, M Müller, T Weiss et al. *Wiley-VCH*, Weinheim, Germany (2004)

97. JAG Agundez: NAT2 genotyping: equilibrium between accuracy and feasibility in routine analyses. *J Appl Res* 3(2), 1-6 (2003)

Abreviations: CI: confidence interval, NAT2: N-acetyltransferase 2, OR: odds ratio, SNP: single nucleotide polymorphism

Key Words: rs1801279 (G191A), rs1041983 (C282T), rs1801280 (T341C), rs1799929 (C481T), rs1799930 (G590A), rs1208 (A803G), rs1799931 (G857A)

Send correspondence to: Silvia Selinski, Ardeystr. 67, 44129 Dortmund, Germany, Tel: 49-231-1084-216, Fax: 49-231-1084-343, E-mail: selinski@ifado.de