Published in final edited form as:
Pharmacogenet Genomics. 2012 April ; 22(4): 322–324. doi:10.1097/FPC.0b013e3283510a51.

Novel tagging SNP rs1495741 and 2-SNPs (rs1041983 and rs1801280) yield high prediction of NAT2 genotype in HapMap samples

Yi Jing He[a], **Michael H. Shapero**[b], and **Howard L. Mcleod**[a,*]

[a]UNC Institute for Pharmacogenomics and Individualized Therapy, University of North Carolina, Chapel Hill, NC

[b]Affymetrix, Inc. 3420 Central Expressway; Santa Clara, CA 95051, USA

Keywords

N-acetyltransferase 2 (NAT2); HapMap; tagSNP; rs1495741; DMET Plus microarray

Both single and two SNP strategies for predicting NAT2 phenotype have been recently proposed [1]. The potential for practical utility caused an assessment of predictive potential across multiple world populations, with the goal of determining predictive ability for global health studies.

A NAT2 genotype profile of 16 SNPs (including rs1041983 and rs1801280) were determined in 595 HapMap samples (59 Utah residents with Northern and Western European ancestry (CEU), 60 Mexican ancestry in Los Angeles, California (MEX), 90 Han Chinese in Beijing, China (CHB), 87 Chinese in Metropolitan Denver, Colorado (CHD), 91 Japanese in Tokyo, Japan (JPT), 89 Luhya in Webuye, Kenya (LWK) and 119 Yoruba in Ibadan, Nigeria (YRI)) using the Affymetrix® DMET™ Plus genotyping platform [2]. The *NAT2* haplotype pairs were then determined by these SNPs via DMET Console 1.0 [2]. According to the Consensus Human Arylamine *N*-Acetyltransferase Gene Nomenclature [3], the phenotype of *NAT2* in HapMap samples was determined based on the haplotype. The genotype calls for tagSNP, rs1495741, was retrieved from the HAPMAP database release#28. Percent of agreement (concordance) and kappa statistics (percent of agreement above and beyond chance alone) were calculated for the genotype of rs1495741, 2-SNP (rs1041983 and rs1801280) and *NAT2* haplotype predicted phenotype. Kappa=0.81 was considered as cutoff to evaluate rs1495741 and 2-SNP panel in this study. The value of area under the curve (AUC) for the receiver operating characteristic (ROC) curve was measured using SPSS Statistics software, version 12.0 (SPSS Inc., Chicago, Illinois, USA). $P<0.05$ was considered as significance in this study.

After merging genotype and haplotype data, totally 476 samples have tagSNP, 2-SNP genotype and haplotype data, thus inferred phenotype. Samples with unknown haplotypes

*Correspondence to Dr. Howard L. Mcleod, UNC Institute for Pharmacogenomics and Individualized Therapy, University of North Carolina – Chapel Hill, Campus Box 7361, Rm 1096 Genetic Medicine Building, 120 Mason Farm Rd, Chapel Hill, NC 27599-7360, USA, Tel: +1-919-966-0512; Fax: +1-919-966-5863; hmcleod@unc.edu.

(n=85, phase ambiguity can cause unknown haplotypes) or no calls in genotypes (n=4) were excluded in this study.

The concordance rate between rs1495741 and the predicted phenotype is 91.4% (kappa=0.86, p<0.0001) (Table 1). This value was superior to the 2-SNP panel (concordance rate 87.0%, kappa=0.80, p<0.001). The AUC value of rs1495741 for the "slow acetylator" population was 0.96 (p<0.001). rs1495741 yield 92% sensitivity and 99% specificity in predicting NAT2 "Slow acetylator" phenotype (Table 2a). However, in 2-SNP panel, this value was 0.85(p<0.001) for "Slow acetylator" phenotype (Table 2b). Similar AUC values were seen for "intermediate and rapid acetylator" prediction with rs1495741 and the 2-SNP panel (Table 2). For specific populations, the discordance rate was 20.7% (kappa=0.63, p<0.0001) and 46% (kappa=0.25, p<0.0001) in Nigerians for rs1495741 and 2-SNP panel respectively (Table 3). In Kenyans, 12.7% (kappa=0.78, p<0.001) were miscalculated by rs1495741 and 30.2% (kappa=0.45, p<0.001) were miscalculated by 2-SNP panel. This value decreased significantly to 1.9%–6.8% (kappa range: 0.89–1, p<0.001) in Asian population (Table 3). Meantime, none of the Caucasian and Mexican populations were miscalculated by either panel.

In the study by Selinski et al, the 2-SNP panel outperformed rs1495741 for higher specificity and lower false discovery rate [1]. However, this was not replicated in our assessment, primarily due to miscalculation of the rapid phenotypes (*4/*13, *12/*13) as intermediate and intermediate phenotypes (*4/*5, *13/*14B, *6A/*13, *7/*13) as slow in 2-SNP panel. When break down our samples by population however, the 2-SNP panel did perform equally as rs1495741 for concordance rate in Caucasian and Mexican populations and better in Asian population. In Nigerians and Kenyans, poor concordance indicated that both tagSNP and 2-SNP panel may not be applicable markers for predicting NAT2 phenotypes in African populations.

References

1. Selinski S, Blaszkewicz M, Lehmann ML, Ovsiannikov D, Moormann O, Guballa C, et al. 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. 2011; 21:673–678. [PubMed: 21750470]
2. Burmester JK, Sedova M, Shapero MH, Mansfield E. DMET microarray technology for pharmacogenomics-based personalized medicine. Methods Mol Biol. 2010; 632:99–124. [PubMed: 20217574]
3. Hein DW, Doll MA, Fretland AJ, Leff MA, Webb SJ, Xiao GH, et al. Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. Cancer Epidemiol Biomarkers Prev. 2000; 9:29–42. [PubMed: 10667461]

Table 1

High concordance between tagSNP, 2-SNP and haplotype predicted *NAT2* phenotype in HapMap population

	Slow		Intermediate		Rapid			Concordance (%)	Kappa (*P* value)
rs1495741	**Frequency**	**Percentage**	**Frequency**	**Percentage**	**Frequency**	**Percentage**	**Total**		
A/A	104	0.92	9	0.08	0	0.00	113		
A/G	3	0.01	218	0.94	11	0.05	232	91.4	0.86 (<0.0001)
G/G	0	0	18	0.	113	0.92	131		
2-SNP									
2+	105	0.71	41	0.28	1	0.01	147		
1	2	0.01	200	0.93	14	0.06	216	87.0%	0.80 (<0.0001)
0	0	0	4	0.04	109	0.96	113		

Statistics: Kappa analysis; 0.61–0.80: substantial agreement; 0.81–1.00: almost perfect agreement.

Pharmacogenet Genomics. Author manuscript; available in PMC 2013 April 1.

Table 2

a. AUC value of tagSNP rs1495741 ROC with respect to *NAT2* phenotype

	AUC of rs1495741 (±SE)	95% CI		*P* value	Sensitivity	Specificity
Slow	0.956±0.015	0.926	0.986	<0.0001	0.92	0.99
Intermediate	0.914±0.015	0.885	0.943	<0.0001	0.94	0.90
Rapid	0.915±0.018	0.879	0.952	<0.0001	0.86	0.97

b. AUC value of 2 SNP ROC with respect to *NAT2* phenotype

	AUC of 2-SNP (±SE)	95% CI		*P* value	Sensitivity	Specificity
Slow	0.854±0.023	0.809	0.899	<0.0001	0.85	0.99
Intermediate	0.876±0.017	0.843	0.910	<0.0001	0.93	0.83
Rapid	0.962±0.012	0.939	0.985	<0.0001	0.97	0.96

Statistic: ROC: receiver operating characteristic; AUC: area under the curve

Table 3

Percentage of disconcordance in HapMap populations

HapMap population	rs1495741 (n)	Kappa (*P* value)	2-SNP (n)	Kappa (*P* value)
CEU	0 (0/40)	1 (<0.0001)	0 (0/40)	1 (<0.0001)
MEX	0 (0/40)	1 (<0.0001)	0 (0/40)	1 (<0.0001)
CHB+CHD	6.8% (11/162)	0.89 (<0.0001)	1.9% (3/162)	0.97 (<0.0001)
JPT	4.8% (4/84)	0.92 (<0.0001)	0 (0/84)	1 (<0.0001)
LWK	12.7% (8/63)	0.78 (<0.0001)	30.2% (19/63)	0.45 (<0.0001)
YRI	20.7% (18/87)	0.63 (<0.0001)	46% (40/87)	0.25 (<0.0001)

CEU: Utah residents with Northern and Western European ancestry from the CEPH collection; MEX: Mexican ancestry in Los Angeles, California; CHB: Han Chinese in Beijing, China; CHD: Chinese in Metropolitan Denver, Colorado; JPT: Japanese in Tokyo, Japan; LWK: Luhya in Webuye, Kenya; YRI: Yoruban in Ibadan, Nigeria

Statistics: Kappa analysis; 0.81–1.00: almost perfect agreement; 0.61–0.80: substantial agreement; 0.41–0.60: Moderate agreement; 0.21–0.40: Fair agreement; 0.0–0.20: Slight agreement