Genome analysis

A web server for inferring the human \( N \)-acetyltransferase-2 (NAT2) enzymatic phenotype from NAT2 genotype

Igor B. Kuznetsov*, Michael McDuffie and Roxana Moslehi*
Gen*NY*Sis Center for Excellence in Cancer Genomics, Department of Epidemiology and Biostatistics, University at Albany, One Discovery Drive, Rensselaer, NY 12144, USA

Received on November 26, 2008; revised on February 5, 2009; accepted on February 26, 2009
Advance Access publication March 4, 2009
Associate Editor: Alfonso Valencia

1 INTRODUCTION

\( N \)-acetyltransferase-2 (NAT2) is an important enzyme that catalyzes the acetylation of aromatic and heterocyclic amine carcinogens. Individuals in human populations are divided into three NAT2 acetylator phenotypes: slow, rapid and intermediate. NAT2PRED is a web server that implements a supervised pattern recognition method to infer NAT2 phenotype from SNPs found in NAT2 gene positions 282, 341, 481, 580, 803 and 857. The web server can be used for a fast determination of NAT2 phenotypes in genetic screens.

Availability: Freely available at http://nat2pred.nl.Albany.edu
Contact: kuznetsov@albany.edu; rmoslehi@albany.edu
Supplementary information: Supplementary data are available at Bioinformatics online.

ABSTRACT

Summary: \( N \)-acetyltransferase-2 (NAT2) is an important enzyme that catalyzes the acetylation of aromatic and heterocyclic amine carcinogens. Individuals in human populations are divided into three NAT2 acetylator phenotypes: slow, rapid and intermediate. NAT2PRED is a web server that implements a supervised pattern recognition method to infer NAT2 phenotype from SNPs found in NAT2 gene positions 282, 341, 481, 580, 803 and 857. The web server can be used for a fast determination of NAT2 phenotypes in genetic screens.

© 2009 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
After the genotype is selected, the user can click 'Submit' button.

### Table 1. The performance of NAT2PRED server

| NAT2 phenotype | Sensitivity (SN) | Specificity (SP) |
|----------------|-----------------|-----------------|
| Rapid          | 99.6% (93.4%)   | 100% (90.1%)    |
| n = 503        | 100%            | 99.7%           |
| Slow           | 100%            | 100%            |
| n = 790        | 92.5% (95.4%)   | 100%            |

The total number of cases for a given phenotype is shown in a corresponding row. SNPs are used, the predictor of the NAT2 acetylator phenotype achieves a nearly perfect accuracy of 99.9% (Equation 1) and nearly perfect class-specific sensitivities and specificities (Equation 2) between 99.6 and 100%. Such a well-balanced performance is observed despite the highly unbalanced nature of the dataset, meaning that the number of subjects with the slow phenotype is almost an order of magnitude larger than that of subjects with the rapid phenotype. Importantly, individuals with the slow phenotype, who are at increased risk of developing tumors, are identified with almost an order of magnitude larger than that of subjects with the rapid phenotype, almost an order of magnitude larger than that of subjects with the rapid phenotype.

### 3 RESULTS

The results of the cross-validation are shown in Table 1. If all six SNPs are used, the predictor of the NAT2 acetylator phenotype is trained using the data on all 1377 subjects. It has a simple intuitive user interface. The user is asked to select a genotype for each of the six SNP loci using radio buttons. The output page displays the selected genotype and the probabilities of each of the three acetylator phenotypes (slow, intermediate and rapid) for these genotypes (Supplementary Fig. 3). The final prediction is the phenotype with the highest probability. There is also an option for a batch submission of genotypes for multiple individuals. Detailed instructions and information about the methodology and output format can be found by clicking the corresponding help hyperlink located on the input page. To the best of the authors’ knowledge, NAT2PRED is the only existing web server for inferring NAT2 acetylator phenotypes from genotyping data. NAT2PRED was developed on a dataset where majority of subjects are Caucasian (94%). However, the prediction model utilizes generally observed linkage disequilibrium between the six NAT2 SNPs and can be applied to individuals from any ethnicity. The web server is publicly available at http://n2apred.rutgers.edu.

### ACKNOWLEDGEMENTS

The authors thank Dr R. B. Hayes, Division of Cancer Epidemiology and Genetics and Drs C. Berg and P. Prorok, Division of Cancer Prevention, NCI, NIH, DHHS, the Screening Center investigators and staff of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, Mr T. Riley and staff, Information Management Services, Inc., Mr B. O’Brien and staff, Westat, Inc. and Drs B. Kopp, W. Shao and staff, SAIC-Frederick. Most importantly, we acknowledge the PLCO study participants for their contributions to making this study possible.

**Conflict of Interest**: none declared.

### REFERENCES

Baldi,P. et al (2000) Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics, 16, 412–424.

Biem,M., et al. (1990) Human arylamine N-acetyltransferase genes: isolation, chromosomal localization, and functional expression. DNA Cell Biol., 9, 193–203.

Brockton,N. et al. (2000) N-acetyltransferase polymorphisms and colorectal cancer: a HaflER Review. Am. J. Epidemiol., 151, 846–861.

C Chang,C.C. and Lin,C.J. (2003) LIBSVM: a library for support vector machines. Available at [http://www.csie.ntu.edu.tw/~cjlin/libsvm](http://www.csie.ntu.edu.tw/~cjlin/libsvm) (last accessed date April 6, 2007).

Hein,D.W. et al. (2000) Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. Cancer Epidemiol. Biomarkers Prev., 9, 29–42.

Hein,D.W. (2006) N-acetyltransferase 2 genetic polymorphism: effects of carcinogen and haplotype on urinary bladder cancer risk. Oncogene, 25, 1649–1658.

Hu,C.W. and Lin,C.J. (2002) A comparison of methods for multi-class support vector machines. IEEE Trans. Neural Netw., 13, 415–425.

Moslehi,R. et al. (2006) Cigarette smoking, N-acetyltransferase genes and the risk of advanced colorectal adenomas. Pharamgenomics, 7, 819–820.

Pottorff,J. et al. (1999) Colorectal adenomatous and hyperplastic polyps: smoking and N-acetyltransferase 2 polymorphisms. Cancer Epidemiol. Biomarkers Prev., 8, 69–75.

Stephens,M. et al. (2001) A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet., 68, 976–989.

Tiemersma,E. (2006) N-acetyltransferase 2 and cigarette smoking. Curr. Opin. Pharmacol., 6, 218–223.

Vapnik,V.N. (1998) Statistical Learning Theory. John Wiley & Sons, New York.