Allele Assignment and haplotype prediction

Allele assignment and haplotype prediction was based on the coding SNPs (cSNPs) 282C>T, 341T>C, 481C>T, 590G>A, 803A>G, 857G>A. 'Unambiguous' alleles were deduced by combinations of the cSNPs with up to one heterozygous cSNP. Haplotypes were inferred using the software package PHASE Version 2.0.2 (URL: http://www.stat.washington.edu/stephens/software.html) [1,2] with inclusion of the positional information of each SNP. Further, haplotypes were estimated with an additional polymorphism in the flanking region (rs2552) which was verified in the SNP500 project, but is not listed in the *NAT* database. Model selection was carried out by the comparison of goodness-of-fit of different models. Expected haplotype frequencies and standard errors for both study populations as well as best haplotype pairs for each individual were calculated. A permutation test for differences in haplotype frequencies between the Caucasian and Kyrgyz study population was performed.

For all cases with more than one heterozygous SNP, many potential alleles could be related. Therefore, we provided the most probable haplotype pairs for each individual, based on the haplotype estimation with six cSNPs. In analogy to the allele nomenclature (prefix asterisk (*)) we used the prefix number (#) to indicate predicted haplotypes. For Romanians, alleles #7B and #11A could be additionally predicted. In Kyrgyz, we found no additional allele.

1 Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001; 68(4):978-989.

2 Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 2003; 73(5):1162-1169.