DOI:10.1158/1055-9965.EPI-06-0759

Cancer Epidemiology, Biomarkers & Prevention

Triallelic Single Nucleotide Polymorphisms and Genotyping Error in Genetic Epidemiology Studies: *MDR1* (*ABCB1*) *G2677/T/A* as an Example

Claudia Hüebner, Ivonne Petermann, Brian L. Browning, et al.

Cancer Epidemiol Biomarkers Prev 2007;16:1185-1192. Published online June 4, 2007.

DOI:10.1158/1055-9965.EPI-06-0759

Triallelic Single Nucleotide Polymorphisms and Genotyping Error in Genetic Epidemiology Studies: *MDR1* (*ABCB1*) *G2677/T/A* as an Example

Claudia Hüebner,[1,4] Ivonne Petermann,[1,4] Brian L. Browning,[1,3,4] Andrew N. Shelling,[2,4] and Lynnette R. Ferguson[1,4]

[1]Discipline of Nutrition and [2]Department of Obstetrics and Gynecology, Faculty of Medical and Health Sciences and [3]Department of Statistics, Faculty of Science, The University of Auckland and [4]Nutrigenomics New Zealand, Auckland, New Zealand

Abstract

Accurate measurement of allele frequencies between population groups with differing sensitivities to disease is fundamental to genetic epidemiology. Genotyping errors can markedly influence the biological conclusions of a study. This issue may be especially important now there is increasing recognition of triallelic single nucleotide polymorphisms (SNPs) in the genome and their possible role in diseases like inflammatory bowel disease. For example, the *MDR1* (*ABCB1*) SNP *G2677/T/A* was, like many other triallelic SNPs, originally described as diallelic. Here, we report a comprehensive analyses of estimated allele frequencies of this SNP in a set of 73 human DNA samples, comparing six commonly used genotyping methods (Applied Biosystems Taqman, Roche LightCycler melting analysis, allelic discrimination PCR, DNA sequencing, Sequenom, and RFLP) from the angle of their error potential. Only Sequenom and DNA sequencing provided accurate measurements, if we had not had prior knowledge of the triallelic nature of this SNP. The other tested methods (with the exception of LightCycler) failed to show any indication of the presence of the rare third *A*- allele in a diallelic assay. Although most of the errors were due to the inability to detect the third allele, all methods except Sequenom and sequencing produced errors for the detection of the two common alleles *G* and *T* (LightCycler, 6 errors; PCR, 4 errors; RFLP, 2 errors; Taqman, 1 error). There is considerable variability in the reported frequencies of the different alleles of the *MDR1 G2677/T/A* SNP, and the role of this SNP in the etiology of inflammatory bowel disease has been controversial. Our data emphasize the importance of choosing the appropriate method for SNP detection and lead us to suggest that part of the previously reported variation may reflect artifacts associated with the different genotyping methodologies used. The failure to recognize the triallic nature of a SNP may lead to underestimations of real genetic associations. (Cancer Epidemiol Biomarkers Prev 2007;16(6):1185–92)

Introduction

The International HapMap Project (1) has opened the door for a new generation of diagnostic tools aimed at identifying and characterizing human diversity. In particular, it has provided a large resource of single nucleotide polymorphisms (SNPs) that provide much of the variation between different individuals and different ethnic groups. Although most of the SNPs associated with human disease have been described as diallelic, in the last few years, an increasing number of these have been recognized to be triallelic and possibly even tetra-allelic. Most of the multiplex techniques that are being increasingly used for genotyping are based on discerning one allele from the other (i.e., start with the assumption that the allele is diallelic; refs. 2, 3). We wished to consider whether starting with such an assumption could impede the discovery of novel triallelic SNPs, and whether alleles may have been mistyped in the past. This would have implications for the accurate estimation of population data.

The group of cancer-prone inflammatory bowel diseases (IBD) includes ulcerative colitis and Crohn's disease. We used the National Center for Biotechnology Information SNP database to identify variants of genes that are described in the literature as associated with IBD susceptibility: *MDR1* (4), *DLG5* (5), *OCTN1/2* (6), *NFkB1* (7), *TNF* and *TNFRSF1B* (8), *MIF* (9), *IL4* (10), and *IL11* (11). Eleven triallelic SNPs have been reported in eight of the identified genes. Interestingly, six of the triallelic SNPs in four of the IBD-associated genes (*MDR1*, *MIF*, *NFkB1*, and *TNFRSF1B*) have been previously described as diallelic (Table 1).

The human *MDR1* gene, located on chromosome 7, encodes an ATP-dependent efflux transporter pump (P-glycoprotein) that is highly expressed in various tissues, including the epithelial surfaces of the intestine. The level of expression of P-glycoprotein is critical in determining the pharmacokinetics of a wide-ranging number of substrates, including anticancer drugs (12-15). There is considerable interindividual variability in P-glycoprotein expression that has implications not only for the development of resistance to various pharmaceutical agents but also for disease susceptibility (16). Several SNPs in the *MDR1* gene have been associated with susceptibility to the development of various types of cancer (16), HIV susceptibility (17), hypercholesteremia (18), and Parkinson's disease (19). They have also, arguably, been associated with IBD (4, 20-24).

The *MDR1* gene is 209 kb in length and composed of 28 exons, and at least 314 SNPs have been described (25-28). Thus far, three variants within the gene (*G2677T/A* in exon 21, *C3435T* in exon 26, and *T129C* in exon 1B) have been shown to correlate with a lower P-glycoprotein expression in normal tissues (26, 29-31). *G2677T* and *C3435T* SNPs are in linkage disequilibrium (multiallelic $D' = 0.85$; refs. 22, 32, 33). Considering the triallelic SNP in exon 21, the reference

Received 9/8/06; revised 2/18/07; accepted 3/30/07.

Grant support: *Nutrigenomics: Tailoring New Zealand foods to people's genes* New Zealand Foundation for Research Science and Technology grant C02X0403.

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.

Note: Nutrigenomics New Zealand is a collaboration among AgResearch Ltd., Crop & Food Research, HortResearch, and The University of Auckland, with funding through the Foundation for Research Science and Technology.

Requests for reprints: Lynnette R. Ferguson, Discipline of Nutrition, Faculty of Medical and Health Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand. Phone: 6493737599, ext. 86372; Fax: 6493035962. E-mail: l.ferguson@auckland.ac.nz

doi:10.1158/1055-9965.EPI-06-0759

DOI:10.1158/1055-9965.EPI-06-0759

Table 1. Known triallelic SNPs in IBD-associated genes as shown on the National Center for Biotechnology Information database

Gene	rs no.	Description as diallelic (year)	Description of 3rd allele (year)
MDR1	rs10274623	*C/G* (2003)	*C/G/T* (2005)
	rs2032582	*G/T* (2001)	*A/G/T* (2003)
TNFRSF1B	rs522205	*A/T* (2000)	*A/T/C* (2003)
OCTN2	rs11568513	—	*A/G/T* (2003)
DLG5	rs1866436	—	*C/G/T* (2001)
IL4	rs2243244	—	*A/G/T* (2001)
IL11	rs4252546	—	*A/C/G* (2002)
MIF	rs2330659	*A/C* (2001)	*A/C/G* (2002)
	rs2330658	*A/T* (2001)	*A/C/G* (2002)
NFkB1	rs12721575	*G/T* (2002)	*A/G/T* (2004)
	rs3810903	—	*A/G/T* (2003)

G2677 is Ala^{893}, with the *T* variant being Ser^{893}, and the less frequent *A* variant coding for Thr^{893}. Various research groups studying IBD have studied SNPs within *MDR1* to determine whether they might be associated with susceptibility to the development of disease. To date, results for the *G2677T/A* polymorphisms have been controversial (4, 20, 21, 23, 24); however, a recent meta-analysis reported evidence for association of the *3435T* allele with ulcerative colitis [odds ratio (OR), 1.12; 95% confidence interval, 1.02-1.23] but not Crohn's disease (22).

We have genotyped DNA samples from a small set of control and Crohn's disease patient samples using a variety of genotyping methods to consider the question as to whether genotyping errors associated with different methods could explain why different studies have not been able to consistently find association to *MDR1* SNPs.

Materials and Methods

Study Population. Seventy-three human subjects were recruited either from the Auckland District Health Board gastroenterology clinics or healthy volunteers to provide approximately equal numbers of male and female subjects and controls or IBD patients. Blood samples were collected into heparinized tubes, and DNA was isolated using the Puregene DNA Purification kit (Gentra Systems) according to the manufacturer's protocol. The amount of DNA extracted was quantified by absorbance spectroscopy (260 and 280 nm) and diluted to 10 ng/μL for working solutions. The isolated DNA was stored at −20°C, and the working solutions were stored at 4°C. The study was conducted under ethical protocol MEC/04/12/011, authorized through the New Zealand Multi-Region Human Ethics Committee.

Genotyping Methods. The PCR, RFLP, and Taqman SNP Genotyping Assay assays were designed to detect a diallelic rather than a triallelic SNP. The allelic discrimination PCR and Taqman SNP Genotyping Assay assays tested for the presence of *G* and *T* alleles, whereas the RFLP detected *G* allele dosage. All primers used for the different assays (except for the primers obtained for Taqman SNP Genotyping Assay) were obtained from Invitrogen. The techniques were done as follows.

PCR for DNA Sequencing or RFLP. Details of the primers used for amplification of exon 21 are provided in Table 2. The sequence of the primers was designed using OligoPerfect Designer free software[5] and checked for specificity using the National Center for Biotechnology Information BLAST server.[6]

The PCR reactions were done in a 25-μL reaction volume containing 20 ng genomic DNA, 100 pmol of each primer, 0.2 mmol/L of each deoxynucleotide triphosphate, 1× PCR buffer, 1.5 mmol/L $MgCl_2$, and 1 unit Taq polymerase (Qiagen). The PCR program for exon 21 consisted of 30 cycles at 94°C for 30 s, 58°C for 30 s, and 72°C for 30 s and a final elongation step at 72°C for 10 min. The PCR products were checked on a 1.5% agarose gel and photographed before being subjected to a RFLP analysis or DNA sequencing.

RFLP Analysis. To determine the respective genotype (*G* or *T*), RFLP analysis with the restriction endonuclease *Bse*YI was conducted after PCR-based amplification (primer listed in Table 2). PCR product (10 μL) was combined with 4 units enzyme, 2 μL of 10× Restriction Enzyme Digestion Buffer 3, and 0.5 μL of bovine serum albumin (all reagents from New England Biolab) in a total volume of 20 μL. Samples were digested for 4 h at 37°C. As the enzyme *Bse*YI remains bound to DNA after digestion and alters migration rate of DNA during electrophoresis, 1 μL of 10% SDS was added after 4 h to disrupt binding. The digestion products were separated on a 2% agarose gel and stained with ethidium bromide.

DNA Sequencing. Amplicons from exon 21 were cleaned according to the manufacturer's instructions using the ChargeSwitch PCR Clean-Up kit (Invitrogen). Automated DNA sequencing was done on an ABI 3130XL Genetic Analyzer sequencer by using BigDye Terminator version 2 reactions (Perkin-Elmer/Applied Biosystems) using the 2677 forward primer.

Conventional Allelic Discrimination PCR. To achieve allelic discrimination between wild-type and mutant allele, two physically separate PCR reactions containing the 2677 forward primer and the corresponding wild-type (2677W) or mutant-specific primer (2677M) were done (Table 2). All reactions were carried out in total volume of 25 μL containing 20 ng genomic DNA, 100 pmol of each primer, 0.2 mmol/L of each deoxynucleotide triphosphate, 1× PCR buffer, 1.5 mmol/L $MgCl_2$, and 1 unit Taq polymerase (Qiagen). The PCR program for allelic discrimination consisted of 30 cycles at 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s and a final elongation step at 72°C for 10 min. The PCR products were electrophoresed on a 1.5% agarose gel, and the genotype assignment was selected on the basis of the following criteria: no visible band represents the absence of the analyzed allele, whereas a band indicates the presence of the analyzed allele.

Applied Biosystems Taqman SNP Genotyping Assay. The SNP at position 2677 of MDR1 was genotyped using the Taqman MGB diallelic discrimination system (34). Probes and oligonucleotides were obtained from Applied Biosystems using the Assay-by-Design product (listed in Table 2). The reactions were prepared by using 2× Taqman Universal Master Mix, 40× SNP Genotyping Assay Mix, DNase-free water, and 10 ng genomic DNA in a final volume of 5 μL per reaction. The PCR amplification was done using the ABI Prism 7900 HT sequence-detector machine under the following conditions: 10 min at 95°C enzyme activation followed by 40 cycles at 92°C for 15 s and 60°C for 1 min (annealing/extension). The allelic discrimination results were determined after the amplification by performing an end-point read.

Roche LightCycler Melting Curve Analysis. The LightCycler combines rapid thermal cycling for PCR with real-time fluorescence monitoring (35, 36). After amplification, the fluorescence signal allows genotyping by analysis of the allele-specific melting behavior of the hybridization probe. The reaction mixture (20 μL) contained 1 unit Taq polymerase, 2 μL of 10× Taq buffer (GeneCraft), 2.5 mmol/L $MgCl_2$, 0.1 mmol/L deoxynucleotide triphosphates (GeneCraft),

[5] http://www.invitrogen.com
[6] http://www.ncbi.nlm.nih.gov/blast/

DOI:10.1158/1055-9965.EPI-06-0759

30 mg/L bovine serum albumin (New England Biolab), 50 ml/L dimethyl sulfoxide (Merck), 0.25 mol/L forward primer, 0.1 mol/L reverse primer, 0.15 mol/L of the anchor, 0.05 mol/L of the locked nucleic acid–modified sensor, 1 μL DNA (40-60 ng/L), and water (PCR grade) up to 20 μL. The following program was done: an initial denaturation at 94°C for 2 min at 20°C/s, followed by a 50-cycle program consisting of heating to 94°C at 20°C/s with no hold, cooling to 58°C at 20°C/s with a 10-s hold, and heating to 72°C at 2°C/s with a 15-s hold. The melting curve was determined by 20 s denaturation at 94°C cooling to 32°C at 20°C/s with a 20 s hold by continuous temperature increase from 32°C to 70°C in increments of 0.1°C/s. Fluorescence was recorded continuously while heating.

Sequenom MassARRAY Genotyping System. Genotyping was carried out with a MassARRAY technique (Sequenom; refs. 37, 38) using a chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (39). Multiplex SNP assays were designed using SpectroDesigner software (Sequenom); 384-well plates containing 2.5 ng DNA in each well were amplified by PCR following the specifications of Sequenom. After PCR, shrimp alkaline phosphatase (Sequenom) was added to samples to prevent future incorporation of unused deoxynucleotide triphosphates that could interfere with the primer extension assay. Allele discrimination reactions were conducted by adding the extension primer(s), DNA polymerase, and a cocktail mixture of deoxynucleotide triphosphates and dideoxynucleotide triphosphates to each well. MassExtend clean resin (Sequenom) was added to the mixture to remove extraneous salts that could interfere with matrix-assisted laser desorption/ionization time-of-flight analysis. Genotypes were determined by spotting an aliquot of each sample onto a 384 SpectroChip (Sequenom), which was subsequently read by the matrix-assisted laser desorption/ionization time-of-flight mass spectrometer. Assay conditions are available upon request and primer sequences are shown in Table 2.

Estimating the Incidence of Triallelic SNPs in Human Populations. The Seattle SNP database (SeattleSNPs, National Heart, Lung, and Blood Institute Program for Genomic Applications, SeattleSNPs, Seattle, WA)[7] was used to estimate the proportion of SNPs that are triallelic. This database contains polymorphisms identified from DNA sequencing 5.9 Mb from 280 genes in a panel of unrelated subjects. The 280 genes were selected because they are thought to influence inflammatory response in humans. Each gene was genotyped in one panel of subjects. The panel of subjects was either a set of 20 individuals of African-American descent and 19 individuals of European descent, or a set of 20 Yoruba from Ibadan, Nigeria from the YRI HapMap panel and 20 Caucasians from Utah from the CEU HapMap panel (1). For simplicity, we refer to European and African samples without distinguishing whether the panel with the African-American samples was used, or the panel with the African samples was used.

Results

Allele Frequencies for *MDR1 G2677/T/A* as Estimated using Six Different Methodologies. The allele frequencies, as estimated by different methods, are shown in Table 3. The true genotype of each sample was defined as the result of matching genotypes of at least four methods. In the case of only three methods with matching results (six samples), Sequenom MassARRAY Genotyping system had to be one of them (note that in all six cases, DNA sequencing agreed with the Sequenom results). In this population, estimates of the proportion of the *G* allele ranged from 0.543 (judged by PCR) to 0.589 (LightCycler). Conversely, the *T* allele appeared lowest when estimated by LightCycler (0.377) and highest using the PCR-based allelic discrimination method (0.457). As most of the known SNPs are diallelic, we wanted to determine what effect the presence of a third allele would have when it is detected in a two-dimensional assay. Thus, the *A/G* genotype appears as a *G/G* genotype and an *A/T* as a *T/T* genotype in the Taqman assay. None of the PCR methods, RFLP, or Taqman SNP Genotyping Assay provided evidence for the presence of the *A* allele. However, although our LightCycler method was not designed using knowledge of the third allele, this allele became obvious from the spectra generated (Fig. 1).

Genotype Errors as Estimated Using Six Different Methodologies. The genotype error analysis is shown in Table 4. Even when the genotype was called incorrectly, one of the two alleles was usually correct. The only exceptions in our data set were for two *T/T* genotypes that the LightCycler incorrectly called as *G/G* and the one *T/T* that RFLP incorrectly called as *G/G*.

Most of the errors (6 of 8 RFLP errors, 6 of 7 Taqman SNP Genotyping Assay errors, and 6 of 10 PCR errors) were due to the inability to detect the *A* allele in the six samples that carried the *A* allele. RFLP incorrectly called two *T/T* genotypes (as *G/T* and *G/G*), and PCR incorrectly called three *G/G* genotypes as *G/T* and one *G/T* genotype as *T/T*. The seven LightCycler errors did not have obvious pattern at the genotypic level or the allelic level. This method called six T alleles incorrectly as *G* alleles and called two *G* alleles and one *A* allele incorrectly as *T* alleles. Even if the six samples with an *A* allele are ignored, allelic discrimination PCR still showed four errors (error rate = 0.063), and RFLP had two errors in 63 non-missing genotypes. Neither DNA sequencing nor Sequenom MassARRAY Genotyping system generated any errors (Table 4).

To exclude the possibility that the *A* allele itself would not be detectable with methods where the knowledge of the third allele is necessary for the assay design, we specifically redesigned two of the assays with respect to the *A* allele (*A/T* Taqman assay and allelic discrimination PCR). Neither the Taqman assay nor allelic discrimination PCR failed to detect this allele (data not shown). However, the two-dimensional nature of the assay design restricts the Taqman assay to be able to detect only two alleles (*A/T* in our case) and results in missing another allele (here the G allele). Accordingly, all samples with a *G/G* genotype failed to amplify, and most of the *G/T* samples were detected as *T/T* or failed to amplify. On the other hand, all *A/T* and *T/T* genotypes were called correctly. However, in the case of an *A/G* genotype, the Taqman assay either calls it as an *A/A* or *A/T* genotype. No method called a *G* or *T* allele as an *A* allele. The six samples that contained an *A* allele were genotyped correctly by the DNA sequencing and Sequenom MassARRAY Genotyping system methods. The LightCycler correctly genotyped five of the six samples containing the *A* allele. However, the number of *A* alleles in our sample was too small to determine the accuracy of these methods when assaying samples carrying the *A* genotype.

Missing Genotype Analysis. It seemed that particular genotypes failed with certain methods (Table 5). Homozygotes (*G/G* or *T/T*) seemed to be preferentially missing when using allelic discrimination PCR and Sequenom, whereas heterozygotes (*G/T*) seemed to be preferentially missing when using DNA Sequencing. The LightCycler and RFLP methods had no missing genotypes.

[7] http://pga.gs.washington.edu, accessed June 25, 2006.

DOI:10.1158/1055-9965.EPI-06-0759

Table 2. Oligonucleotide sequences for primers used for DNA sequencing, RFLP, allelic discrimination PCR, Sequenom, and Taqman SNP Genotyping Assay

Primer	5′ Position	Sequence	3′ Position
Sequencing			
2677Cfor	65436	GCTATAGGTTCCAGGCTTGCT	65416
MDR1rev	65140	TAGAGCATAGTAAGCAGTAGG	65161
RFLP			
MDR1 forward	65304	TGCAATAGCAGGAGTTGT	65287
MDR1 reverse	64964	AAAGTGGGGAGGAAGGAAGA	64983
Allelic discrimination			
2677W	65221	AGTTTGACTCACCTTCCCTGC	65241
2677M	65221	AGTTTGACTCACCTTCCCTGA	65241
Taqman primer			
Forward	65461	GTCTTGGACAAGCACTGAAAGATAAGA	65435
Reverse	65186	CATATTTAGTTTGACTCA	65232
Probe 1	65233	VIC- CTTCCCAGAACCTTC-NFQMGB	65247
Probe 2	65235	FAM- TCCCAGCACCTTC-NFQMGB	65247
LightCycler primer			
MDR1 ex21S forward	65297	GCAGGAGTTGTTGAAATGAAAATG	65274
MDR1 ex21B reverse	65218	cgcctgc TTTAGTTTGACTCA	65232
21 Anchor	65253	CTTTCTTATCTTTCAGTGCTTGTCC	65276
21 Sensor	65248	TTCCCAGTACCTTCT	65235
Sequenom primer			
MDR1 PCR forward	65290	ACGTTGGATGGAAAATGTTGTCTGGACAAGC	65270
MDR1 PCR reverse	65214	ACGTTGGATGCATATTTAGTTTGACTCACC	65233
MDR1 UEP_SEQ	65262	ggcGATAAGAAAGAACTAGAAGGT	65240
MDR1 EXT1_SEQ	65262	ggcGATAAGAAAGAACTAGAAGGTC	65241
MDR1 EXT2_SEQ	65262	ggcGATAAGAAAGAACTAGAAGGTA	65241
MDR1 EXT3_SEQ	65262	ggcGATAAGAAAGAACTAGAAGGTG	65241
MDR1 EXT4_SEQ	65262	ggcGATAAGAAAGAACTAGAAGGTT	65241

NOTE: Primers were designed on the published MDR1 sequence (AC005068) or adopted from Song et al. (53).
Abbreviations: NFQ-MGB, non-fluorescent quencher/minor groove binder; VIC, fluorescent dye used to label the Taqman SNP Genotyping Assay probe that detects the allele 1 sequence; FAM, fluorescent dye used to label the Taqman SNP Genotyping Assay probe that detects the allele 2 sequence; UEP, unextended primer; EXT1, EXT2, EXT3, EXT4, mass extent primer.

Estimation of Population Frequencies of Triallelic SNPs. The Seattle SNP database contained 29,827 diallelic SNPs, 67 triallelic SNPs, and 2,070 insertion/deletion polymorphisms. Therefore, 0.224% of the SNPs in the Seattle SNPs database are triallelic. Of the 67 triallelic SNPs, 12 were triallelic in the European samples, and 53 were triallelic in the African samples. Ten of the SNPs were diallelic in the European samples and in the African samples but were triallelic in the combined samples because the European and African samples had different minor alleles. In the African samples, 19 triallelic SNPs had all three allele frequencies >0.05, and in the European samples, five triallelic SNPs had all three allele frequencies >0.05.

Discussion

It is recognized that some techniques (DNA sequencing and Sequenom MassARRAY Genotyping system analysis) can detect a third allele without knowing of its existence. Our data set suggested this was also true, at least for this allele at this locus, for the LightCycler method. However, most of the multiplex techniques that are being increasingly used for genotyping start with the assumption that the SNP is diallelic (Taqman SNP Genotyping Assay and allelic discrimination PCR) and would need the knowledge of a third allele being present for the assay design, although in some cases, a third allele can be detected by examination of the raw data before analysis (40). This is also true for RFLP, which is still commonly used for genotyping. Thus, our assay designs for genotyping analysis were based on the assumption that there are only two alleles (*G* and *T*), ignoring the presence of the rare *A* allele. As anticipated, several of the methods failed to provide signals that would have led us to suspect a third allele. Unexpectedly, however, it was not only the *A* allele that provided difficulties in genotyping with some of the tested methodologies.

Other than hypothesized, it was not apparent that any of the different detection techniques favored one allele over the other. Among all methods tested, the LightCycler and RFLP methods were the only methods that showed no unclear or failed results. Five of seven RFLP genotype errors, five of the six Taqman SNP Genotyping Assay errors, and five of the eight allelic discrimination PCR errors were due to the inability to detect the *A* allele, as would be expected for these methods. To analyze this further, we designed two sets of assays (Taqman and allelic discrimination PCR) to detect the *A* allele and have rerun these new assays through our sample set. Both Taqman assay and allelic discrimination PCR provide accurate measurement for the rare *A* allele.

For family-based studies, genotype error can be a serious problem because it can increase the false positive proportion (41). For case control studies, genotype error generally will cause a loss in power to detect marker-disease associations but not an increase in the false-positive proportion.

Table 3. The observed genotype frequencies obtained using the different methods of analysis (see Materials and Methods)

Method (no. genotypes)	Estimated allele frequency		
	G	T	A
PCR (69)	0.543	0.457	0.000
Taqman (71)	0.586	0.414	0.000
RFLP (73)	0.582	0.418	0.000
LightCycler (73)	0.589	0.377	0.034
Sequencing (69)	0.572	0.384	0.043
Sequenom (69)	0.565	0.391	0.043
True genotype frequency (73)	0.562	0.397	0.041

NOTE: The number of genotypes is <73 for some platforms due to failed or unclear assays (see Table 4).

DOI:10.1158/1055-9965.EPI-06-0759

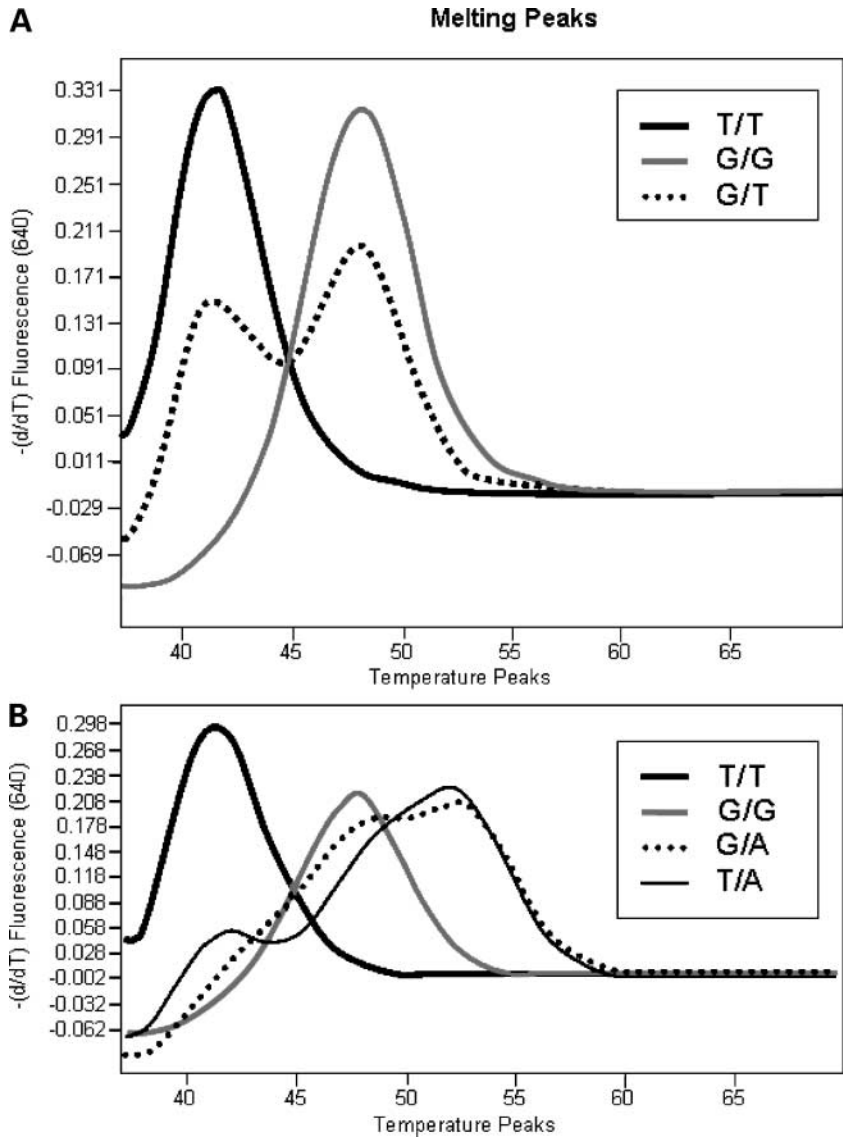

Figure 1. *MDR1 G2677T/A* allelic discrimination PCR by melting curve analysis using the LightCycler. **A.** Three common genotypes: *T/T, G/T*, and *G/G*. **B.** Genotypes *T/T* and *G/G* and the rare genotypes *T/A* or *G/A*, respectively.

The consequences of not detecting a null (i.e., unknown) allele of a triallelic SNP are serious when the null allele affects risk of disease. This can be shown by calculating the population OR for each of the detected alleles in cases and controls for a population that is in Hardy-Weinberg equilibrium. We assume that there are three alleles (a, b, and c) with frequencies p_a, p_b, and p_c, and that allele c is a null allele such that null allele homozygotes have missing genotypes and null allele heterozygotes are miscalled as homozygote genotypes for the non-null allele. Denote the population disease prevalence by ϕ and the genotypic relative risk for c allele heterozygotes as r. Then the apparent population allelic OR for the a allele is:

$$(p_a + rp_c + (0.5)p_b)(p_b + sp_c + (0.5)p_a)/$$
$$(p_a + sp_c + (0.5)p_b)(p_b + rp_c + (0.5)p_a)$$

where $s = (1 - \phi r) / (1 - \phi)$.

For example, when the disease prevalence is <0.1, and the risk allele of a triallelic SNP has an allelic OR of 3.0, if the risk allele is not detected, the allelic OR for each of the detected

Table 4. Error rates for Taqman, LightCycler, PCR, and RFLP

Method	Genotype error (95% confidence interval)	Allele error	Comments
Knowledge of 3rd allele not necessary			
Sequencing	0.014 (0.000, 0.078)	0.007 (0.000, 0.040)	1 error, 68 correct, 4 missing
Sequenome	0.000 (0.000, 0.052)	0.000 (0.000, 0.026)	0 errors, 69 correct, 4 missing
LightCycler	0.096 (0.039, 0.188)	0.062 (0.029, 0.114)	7 errors, 66 correct, 0 missing
Knowledge of 3rd allele necessary			
Taqman	0.010 (0.041, 0.195)	0.050 (0.020, 0.100)	7 errors, 63 correct, 3 missing
PCR	0.145 (0.072, 0.250)	0.072 (0.035, 0.129)	10 errors, 59 correct, 4 missing
RFLP	0.110 (0.049, 0.205)	0.062 (0.029, 0.114)	8 errors, 65 correct, 0 missing

NOTE: Genotype error rate was defined as the number of correct genotypes divided by the number of successful genotypes. Allele error rate is defined as number of correctly called alleles divided by twice the number of successful genotypes. A successful genotype is a genotype that is not missing and that is not unclear.

DOI:10.1158/1055-9965.EPI-06-0759

Table 5. Missing genotypes out of 73 attempted

Method	Missing rate (95% confidence interval)	Comments	Missing alleles
Knowledge of 3rd allele not necessary			
Sequencing	0.055 (0.015, 0.134)	4 missing	*G/T* (4)
Sequenome	0.055 (0.015, 0.134)	4 missing	*G/G* (2), *T/T* (2)
LightCycler	0.000 (0.000, 0.049)	0 missing	
Knowledge of 3rd allele necessary			
Taqman	0.041 (0.009, 0.115)	3 missing	*G/T* (2), *T/T* (1)
PCR	0.055 (0.015, 0.134)	4 missing	*G/T* (2), *G/G* (2)
RFLP	0.000 (0.000, 0.049)	0 missing	

NOTE: Failed and unclear genotypes counts were combined to obtain missing genotype counts.

alleles will be <1.25. In the worst case, when the disease model is recessive, or when the detected alleles have equal population frequency, the allelic OR for each of the detected alleles will appear as 1.0, and there will be no power to detect the true disease association.

When the null allele has low frequency, and there is a sufficiently small difference in the frequencies of the detected alleles, the presence of a null allele is unlikely to cause high levels of missing data or departures from Hardy-Weinberg equilibrium. In cases where a null allele does contribute to an unacceptably high level of missing data or departure from Hardy-Weinberg equilibrium, this evidence of error in the genotyping assay will often result in the SNP being dropped from the analysis.

If an assay (like the Taqman SNP Genotyping Assay) is not designed to detect an allele of a triallelic SNP, there will be little or no power to detect an association of the undetected allele with a disease. Even if all three alleles of a triallelic SNP are detected, high error rates, such as those observed with the LightCycler, can cause substantial loss in power (42).

It was apparent that particular genotypes failed with certain methods. Homozygotes (G/G and T/T) seemed to be preferentially missing when using allelic discrimination PCR and Sequenom MassARRAY Genotyping system, whereas

Table 6. Summary of *MDR1 G2677/T/A* allele frequencies reported in various studies

Study	*n*	Racial group	Methods	Allele frequencies		
				G	T	A
Healthy population						
Cascorbi et al. 2001	461 (167 ♀, 294 ♂)	German	RFLP and sequencing	0.564	0.416	0.019
Kurzawski et al. 2006	204 (♀ 93, ♂ 111)	Polish-Caucasian	Allele-specific PCR and sequencing	0.595	0.385	0.020
Gaikovitch et al. 2003	290 (healthy or non malignant disease)	Russian-Caucasian	Hybridization probe	0.548	0.419	0.033
Allabi et al. 2005	111	West African (Beninese)	2× sequencing	0.991	0.009	0
Cavaco et al. 2003	100	Caucasian Portuguese	PCR-RFLP	0.525	0.475	—
Tan et al. 2004	104	Chinese	Sequencing	0.505	0.437	0.058
	139	Polish Caucasians		0.576	0.414	0.011
Tang et al. 2002	104	Chinese	RFLP	0.505	0.437	0.058
	93	Malay		0.575	0.360	0.065
	68	Indian		0.338	0.618	0.044
Lee et al. 2005	632	Koreans	Pyrosequencing	0.438	0.391	0.171
	142	Vietnamese		0.581	0.356	0.063
Horinouchi et al. 2002	117	Japanese	PCR-RFLP and sequencing	0.440	0.360	0.200
Saito et al. 2003	130 (♀ 70, ♂ 60)	Japanese	Taqman and direct sequencing?	0.432	0.408	0.169
IBD studies						
Urcelay et al. 2006	321 CD	Spanish	PCR and sequencing	0.632	0.359	0.009
	330 UC			0.628	0.365	0.007
	352 controls			0.605	0.384	0.011
Potocnik et al. 2004	139 CD	Slovenian	Taqman	0.595	0.405	—
	144 UC			0.520	0.480	—
	355 controls			0.597	0.403	—
Onnie et al. 2006	828 CD	British (Jewish and non-Jewish)	Pyrosequencing and partial sequencing (48 samples)	0.579	0.405	0.016
				0.533	0.446	0.021
	580 UC			0.579	0.396	0.025
	285 controls					
Ho et al. 2005	335 UC patients	Scottish	Taqman (only *G/T*) and sequencing (100) for A allele frequency	0.546	0.498	0.02
	268 CD patients			0.528	0.472	—
	370 controls			0.512	0.498	—
Palmieri et al. 2005	478 CD patients	Italian	Sequencing	0.559	0.425	0.016
	468 UC patients			0.528	0.447	0.025
	450 controls			0.556	0.423	0.021
Brant et al. 2003	211 non-Jewish IBD	65% non-Jewish		0.602 NJP	0.393 NJP	0.005 NJP
	392 non-Jewish controls			0.524 NJC	0.45 NJC	0.026 NJC
	114 Jewish IBD		PCR-sequencing and pyrosequencing	0.627 JP	0.368 JP	0.005 JP
	219 White Ashkenazi Jewish controls	35% Jewish ancestry		0.605 JC	0.365 JC	0.003 JC

NOTE: ♀, male; ♂, female.
Abbreviations: CD, Crohn's disease; UC, ulcerative colitis; NJP, non-Jewish IBD patients; NJC, non-Jewish controls; JP, Jewish IBD patients; JC, white Ashkenazi Jewish controls.

heterozygotes (*G/T*) seemed to be preferentially missing when using DNA sequencing. The genotyping failures of all methods are not based on the DNA quality, as no sample failed with more then one method.

Another explanation for this failure rate could be the occurrence of allelic dropouts, whereby an unknown polymorphism exists on the template DNA strand where the PCR primer anneals (43, 44). This is unlikely to explain our results, as all our samples were previously sequenced over the whole area of primer annealing and did not reveal any further unknown polymorphisms. We have reviewed the SNP database and are unable to find any reported SNPs within the design of the primers. Although we were unable to obtain the sequence of the area spanning the forward sequencing primer binding site and possible/linked variants, we note that we have successfully used this primer for the allelic discrimination PCR, and we have never found the same samples failing with both methods. This makes it highly unlikely that there are other SNPs in that region. Although we cannot exclude the possibility of introduced errors during the process of primer synthesis, which might lead to the occurrence of null alleles as a consequence of inefficient amplification due to primer/template mismatches (43, 45), we consider that this is unlikely.

Non-random patterns of missing genotypes introduce noise into case-control studies but can cause apparent overtransmission to affected offspring in family-based studies even if the polymorphism is not associated with the disease (46). Our sample sizes were too small to give strong evidence that any of the genotyping platforms gave non-random patterns of missing genotypes. We note that 54.4% of the genotypes in our sample were homozygotes; yet, all four missing genotypes for the Sequenom MassARRAY Genotyping system platform were homozygotes, and all four of the missing genotypes from DNA sequencing were heterozygotes.

For case-control studies, genotype error and non-random missing genotypes can also inflate type 1 error above the nominal rate when using allelic tests that assume Hardy-Weinberg equilibrium (47), such as the χ^2 test or Fisher's exact test.

The distribution of SNPs at the *MDR1 G2677T/A* locus (rs2032582) has been reported to vary across population groups and has shown variable association with IBD. We have summarized published information on ethnic variations in unselected populations and reports on IBD patients (Table 6). Although Schwab et al. (48) suggested that the Ser^{893} variant increased susceptibility in ulcerative colitis but not Crohn's disease, Brant et al. (20) suggested that the reference genotype (*G2677*) increased risk, whereas other studies have failed to show an association. The allelic frequency of the *A* variant was reported to range from 4.4% to 21% in Asians (49, 50) compared with 0.7% to 10% in White subjects (24, 51) and 0.5% in Black subjects (52). It is noteworthy that several different techniques have been used across different laboratories. From our data, comparisons across studies using different methodologies could be considerably misleading.

Some of the techniques that we have used, such as the Taqman SNP Genotyping Assay and RFLP, were designed on the assumption that the target was diallelic. We are aware that these assays can be re-designed to accommodate additional alleles; however, this would add to the cost of the assays and also the time involved in analyzing the samples. For RFLP, it would also depend on the presence of a suitable restriction enzyme. Clearly, when considering the type of genotyping platform to be used in a large association study, along with considering such things as the cost of genotyping, ability to multiplex, the time and handling involved, and/or access to the technology, one also needs to consider the likelihood of encountering a multiallelic SNP in the collection of SNPs being analyzed.

On the basis of the triallelic variant *G2677T/A* in the *MDR1* gene, we have shown different detection techniques (Taqman SNP Genotyping Assay, LightCycler, allelic discrimination PCR, DNA sequencing, Sequenom MassARRAY Genotyping system, and RFLP). Our data lead us to suggest that multiallelic SNPs may be more common than generally realized, may have been overlooked in some studies, and could lead to erroneous overestimation of the frequency of certain alleles. In general, we conclude that more attention is required in the initial analysis of SNPs to determine whether they are multiallelic, perhaps by DNA sequence analysis of a reasonable number of samples. We consider that in some situations, failure to recognize the triallelic nature of the SNP may lead to the over or underestimation of real genetic associations.

References

1. Consortium TIH. A haplotype map of the human genome. Nature 2005;437:1299–320.
2. Wijsman EM, Daw EW, Yu CE, et al. Evidence for a novel late-onset Alzheimer disease locus on chromosome 19p13.2. Am J Hum Genet 2004;75:398–409.
3. Huang YT, Zhang K, Chen T, Chao KM. Selecting additional tag SNPs for tolerating missing data in genotyping. BMC Bioinformatics 2005;6:263.
4. Potocnik U, Ferkolj I, Glavac D, Dean M. Polymorphisms in multidrug resistance 1 (MDR1) gene are associated with refractory Crohn disease and ulcerative colitis. Genes Immun 2004;5:530–9.
5. Stoll M, Corneliussen B, Costello CM, et al. Genetic variation in DLG5 is associated with inflammatory bowel disease. Nat Genet 2004;36:476–80.
6. Waller S, Tremelling M, Bredin F, et al. Evidence for association of OCTN genes and IBD5 with ulcerative colitis. Gut 2006;55:809–14.
7. Karban AS, Okazaki T, Panhuysen CI, et al. Functional annotation of a novel NFKB1 promoter polymorphism that increases risk for ulcerative colitis. Hum Mol Genet 2004;13:35–45.
8. Sashio H, Tamura K, Ito R, et al. Polymorphisms of the TNF gene and the TNF receptor superfamily member 1B gene are associated with susceptibility to ulcerative colitis and Crohn's disease, respectively. Immunogenetics 2002;53:1020–7.
9. Nohara H, Okayama N, Inoue N, et al. Association of the -173 G/C polymorphism of the macrophage migration inhibitory factor gene with ulcerative colitis. J Gastroenterol 2004;39:242–6.
10. Klein W, Tromm A, Griga T, et al. Interleukin-4 and interleukin-4 receptor gene polymorphisms in inflammatory bowel diseases. Genes Immun 2001;2:287–9.
11. Klein W, Rohde G, Arinir U, et al. A promotor polymorphism in the Interleukin 11 gene is associated with chronic obstructive pulmonary disease. Electrophoresis 2004;25:804–8.
12. Thorgeirsson SS, Silverman JA, Gant TW, Marino PA. Multidrug resistance gene family and chemical carcinogens. Pharmacol Ther 1991;49:283–92.
13. Gupta KP, Ward NE, Gravitt KR, Bergman PJ, O'Brian CA. Partial reversal of multidrug resistance in human breast cancer cells by an N-myristoylated protein kinase C-alpha pseudosubstrate peptide. J Biol Chem 1996;271:2102–11.
14. Hsia TC, Lin CC, Wang JJ, Ho ST, Kao A. Relationship between chemotherapy response of small cell lung cancer and P-glycoprotein or multidrug resistance-related protein expression. Lung 2002;180:173–9.
15. Burger H, Foekens JA, Look MP, et al. RNA expression of breast cancer resistance protein, lung resistance-related protein, multidrug resistance-associated proteins 1 and 2, and multidrug resistance gene 1 in breast cancer: correlation with chemotherapeutic response. Clin Cancer Res 2003;9:827–36.
16. Ferguson LR, De Flora S. Multiple drug resistance, antimutagenesis and anticarcinogenesis. Mutat Res 2005;591:24–33.
17. Verstuyft C, Marcellin F, Morand-Joubert L, et al. Absence of association between MDR1 genetic polymorphisms, indinavir pharmacokinetics and response to highly active antiretroviral therapy. AIDS 2005;19:2127–31.
18. Kajinami K, Brousseau ME, Ordovas JM, Schaefer EJ. Polymorphisms in the multidrug resistance-1 (MDR1) gene influence the response to atorvastatin treatment in a gender-specific manner. Am J Cardiol 2004;93:1046–50.
19. Furuno T, Landi MT, Ceroni M, et al. Expression polymorphism of the blood-brain barrier component P-glycoprotein (MDR1) in relation to Parkinson's disease. Pharmacogenetics 2002;12:529–34.
20. Brant SR, Panhuysen CI, Nicolae D, et al. MDR1 Ala893 polymorphism is associated with inflammatory bowel disease. Am J Hum Genet 2003;73:1282–92.
21. Ho GT, Nimmo ER, Tenesa A, et al. Allelic variations of the multidrug resistance gene determine susceptibility and disease behavior in ulcerative colitis. Gastroenterology 2005;128:288–96.
22. Onnie CM, Fisher SA, Pattni R, et al. Associations of allelic variants of the multidrug resistance gene (ABCB1 or MDR1) and inflammatory bowel disease and their effects on disease behavior: a case-control and meta-analysis study. Inflamm Bowel Dis 2006;12:263–71.

DOI:10.1158/1055-9965.EPI-06-0759

23. Palmieri O, Latiano A, Valvano R, et al. Multidrug resistance 1 gene polymorphisms are not associated with inflammatory bowel disease and response to therapy in Italian patients. Aliment Pharmacol Ther 2005;22:1129–38.
24. Urcelay E, Mendoza JL, Martin MC, et al. MDR1 gene: susceptibility in Spanish Crohn's disease and ulcerative colitis patients. Inflamm Bowel Dis 2006;12:33–7.
25. Birney E, Andrews D, Caccamo M, et al. Ensembl 2006. Nucleic Acids Res 2006;34:D556–61.
26. Hoffmeyer S, Burk O, von Richter O, et al. Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity *in vivo*. Proc Natl Acad Sci U S A 2000;97:3473–8.
27. Cascorbi I, Gerloff T, Johne A, et al. Frequency of single nucleotide polymorphisms in the P-glycoprotein drug transporter MDR1 gene in white subjects. Clin Pharmacol Ther 2001;69:169–74.
28. Kim RB, Leake BF, Choo EF, et al. Identification of functionally variant MDR1 alleles among European Americans and African Americans. Clin Pharmacol Ther 2001;70:189–99.
29. Ameyaw MM, Regateiro F, Li T, et al. MDR1 pharmacogenetics: frequency of the C3435T mutation in exon 26 is significantly influenced by ethnicity. Pharmacogenetics 2001;11:217–21.
30. Slovak ML, Kopecky KJ, Wolman SR, et al. Cytogenetic correlation with disease status and treatment outcome in advanced stage leukemia post bone marrow transplantation: a Southwest Oncology Group study (SWOG-8612). Leuk Res 1995;19:381–8.
31. Tanabe M, Ieiri I, Nagata N, et al. Expression of P-glycoprotein in human placenta: relation to genetic polymorphism of the multidrug resistance (MDR)-1 gene. J Pharmacol Exp Ther 2001;297:1137–43.
32. Lewontin RC. The interaction of selection and linkage. I. General considerations;heterotic models. Genetics 1964;49:49–67.
33. Hedrick PW. Gametic disequilibrium measures: proceed with caution. Genetics 1987;117:331–41.
34. Livak KJ. Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. Genet Anal 1999;14:143–9.
35. Arjomand-Nahad F, Diefenbach K, Landt O, Gaikovitch E, Roots I. Genotyping of the triallelic variant G2677T/A in MDR1 using LightCycler with locked-nucleic-acid-modified hybridization probes. Anal Biochem 2004;334:201–3.
36. Pals G, Pindolia K, Worsham MJ. A rapid and sensitive approach to mutation detection using real-time polymerase chain reaction and melting curve analyses, using BRCA1 as an example. Mol Diagn 1999;4:241–6.
37. Leushner J, Chiu NH. Automated mass spectrometry: a revolutionary technology for clinical diagnostics. Mol Diagn 2000;5:341–8.
38. Jurinke C, Denissenko MF, Oeth P, et al. A single nucleotide polymorphism based approach for the identification and characterization of gene expression modulation using MassARRAY. Mutat Res 2005;573:83–95.
39. Bray MS, Boerwinkle E, Doris PA. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. Hum Mutat 2001;17:296–304.
40. Carlson CS, Smith JD, Stanaway IB, Rieder MJ, Nickerson DA. Direct detection of null alleles in SNP genotyping data. Hum Mol Genet 2006;15:1931–7.
41. Mitchell AA, Cutler DJ, Chakravarti A. Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test. Am J Hum Genet 2003;72:598–610.
42. Gordon D, Finch SJ, Nothnagel M, Ott J. Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. Hum Hered 2002;54:22–33.
43. Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors: causes, consequences and solutions. Nat Rev Genet 2005;6:847–59.
44. Walsh PS, Erlich HA, Higuchi R. Preferential PCR amplification of alleles: mechanisms and solutions. PCR Methods Appl 1992;1:241–50.
45. Paetkau D, Strobeck C. The molecular basis and evolutionary history of a microsatellite null allele in bears. Mol Ecol 1995;4:519–20.
46. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95–108.
47. Sasieni PD. From genotypes to genes: doubling the sample size. Biometrics 1997;53:1253–61.
48. Schwab M, Schaeffeler E, Marx C, et al. Association between the C3435T MDR1 gene polymorphism and susceptibility for ulcerative colitis. Gastroenterology 2003;124:26–33.
49. Tang K, Ngoi SM, Gwee PC, et al. Distinct haplotype profiles and strong linkage disequilibrium at the MDR1 multidrug transporter gene locus in three ethnic Asian populations. Pharmacogenetics 2002;12:437–50.
50. Komoto C, Nakamura T, Sakaeda T, et al. MDR1 haplotype frequencies in Japanese and Caucasian, and in Japanese patients with colorectal cancer and esophageal cancer. Drug Metab Pharmacokinet 2006;21:126–32.
51. Gerloff T, Schaefer M, Johne A, et al. MDR1 genotypes do not influence the absorption of a single oral dose of 1 mg digoxin in healthy white males. Br J Clin Pharmacol 2002;54:610–6.
52. Kroetz DL, Pauli-Magnus C, Hodges LM, et al. Sequence diversity and haplotype structure in the human ABCB1 (MDR1, multidrug resistance transporter) gene. Pharmacogenetics 2003;13:481–94.
53. Song P, Li S, Meibohm B, et al. Detection of MDR1 single nucleotide polymorphisms C3435T and G2677T using real-time polymerase chain reaction: MDR1 single nucleotide polymorphism genotyping assay. AAPS PharmSci 2002;4:E29.