Conflicting results in SNP genotype assessment

Lise Lotte Hansen[1], Bo Eskerod Madsen[2], Kristina Pedersen,[1] and Carsten Wiuf[2]

[1]*Danish Centre for Molecular Gerontology, Institute of Human Genetics, University of Aarhus, Denmark and* [2]*Bioinformatics Research Center (BiRC), University of Aarhus, Denmark*

BioTechniques 43:756-762 (December 2007)
doi 10.2144/000112675

Single nucleotide polymorphisms (SNPs) are highly abundant in the genome and especially useful in the search for disease susceptibility genes via population-based association or linkage studies. Therefore, there is a strong need for high throughput and reliable methodologies to assess the SNP genotypes. Despite an unambiguous result of an SNP analysis, with the use of a commercial kit based on primer extension, subsequent sequencing analysis revealed that a proportion of the genotypes was not correctly assessed. The problem we have encountered may originate from specific structures in the genomic DNA sequence, rather than being a methodological problem.

Single nucleotide polymorphism (SNP) is the most common variation of the human genome with more than 9 million reported in public databases (1,2). Extensive studies are being made to characterize single SNPs, or haplotypes containing multiple SNPs, which can be used as prognostic and predictive disease markers or as tools to locate new disease genes.

A number of techniques to establish the genotype of a known SNP are available, as reviewed in Reference 3. The choice of methodology is dependent on available technology platforms and whether the analysis demands a high or low throughput.

The SNaPshot Multiplex kit from Applied Biosystems (Foster City, CA, USA), based upon the primer extension principle, is widely used to assess multiple genotypes in population-based analyses. The output is in the range of 1640–41,280 genotypes per day and depends on the available type of electrophoresis instrument and type of polymer. Genetic analyzer instruments and analysis software are available to most laboratories, which is one reason why the SNaPshot method is straightforward and affordable to apply. The final results are fast and easy to interpret via the GenoTyper or GeneMapper software (Applied Biosystems), which further allows post-processing of the data with other software tools.

In the search for new cancer susceptibility genes and risk markers, we have established the SNP genotypes in different target genes. The SNaPshot Multiplex kit was used to assess the genotypes. When the results were confirmed by sequencing, we discovered that certain SNPs were not correctly interpreted, despite an unambiguous SNaPshot result.

The following procedure was used: Included in the study were 199 patients diagnosed with primary breast cancer and 512 healthy Danish medical students. The group of patients is described in Reference 4. DNA was extracted from peripheral blood and purified by a modified salt precipitation procedure from each individual (5).

Sixteen SNPs were genotyped spanning the RNase L (*RNASEL*) and the regulator of G-protein signaling 16 (*RGS16*) genes, as well as seven SNPs within the 2′–5′ oligoadenylate synthetase 1 (*OAS1*) gene. The protocol given below is general, but the annealing temperatures are specific for the SNPs: rs3738579 (*RNASEL*), rs2660 (*OAS1*), rs1051042 (*OAS1*), and rs10774671 (*OAS1*).

The manufacturer's protocol was followed. In principle, a primer is designed to anneal to the target DNA just 5′ to the polymorphic nucleotide. The primers can complement either the forward or the reverse DNA strand. In a multiplex reaction a tail of tetramers (GACT) is linked to the 5′ end to ensure different lengths of the primers and separation during electrophoresis.

The criteria used for SNP primer design was in accordance with the SNaPshot protocol: no hairpin structures, no self- or cross-annealing, ~50% GC content, T_m between 55° and 60°C, and 18–22 bp in length excluding the GACT tail. The location of the primer is restricted to the area flanking the SNP on either of the two DNA strands.

The test DNA fragments, containing one or multiple SNPs, are PCR-amplified using 20 ng of genomic DNA, 1 × NH4 buffer (supplied with the Taq polymerase), 5 pmol of each primer (10 pmol for amplification of the *OAS1* fragment), 250 µmol dNTP, 0.5 U Taq polymerase (Ampliqon, Bie and Berntsen, Herlev, Denmark) and 1 M betaine (*OAS1*) in a final volume of 20 µl. Primers are described in Table 1.

The PCR amplification of *RNASEL* exon 2 (rs3738579) consisted of an initial denaturing step of: 40 s at 96°C, followed by 34 cycles comprising 35 s at 93°C, 40 s at 56°C and 40 s at 72°C, and terminated by 6 min at 72°C. PCR amplification of fragments from *OAS1* exons 3 and 7 consisted of an initial denaturing step for 1 min at 95°C followed by 32 cycles comprising 30 s at 95°C, 30 s at 60°C and 90 s at 72°C, and terminated by 7 min at 72°C.

The amplification products were treated enzymatically to remove unincorporated deoxynucleotides (dNTPs) and primers by adding 3.3 U shrimp alkaline phosphatase (SAP; Roche Diagnostics, Basel, Switzerland) and 1.34 U exonuclease 1 (Exo1; Medinova, Glostrup, Denmark) to a 10-µl PCR product, and incubating it for 75 min at 37°C followed by 15 min at 75°C to inactivate the enzymes.

For the primer extension, 2 pmol of the specific SNP primer and 3.0 µl SNaPshot mix were added to 6 µl of the enzymatically-treated PCR product in a final volume of 10 µl. Primer extension was performed in 25 cycles comprising 10 s at 96°C, 5 s at 50°C, and 30 s at 60°C. The product was left at 4°C for further processing by adding 1 U SAP and incubating at 37°C for 75 min followed by 15 min at 75°C.

No dNTPs were present in the extension reaction, only dideoxynucleotides (ddNTPs), each labeled with a unique fluorescent dye. The primer was extended by one nucleotide complementary to the SNP allele. After primer extension, the excess ddNTPs were removed by enzymatic treatment. The extended

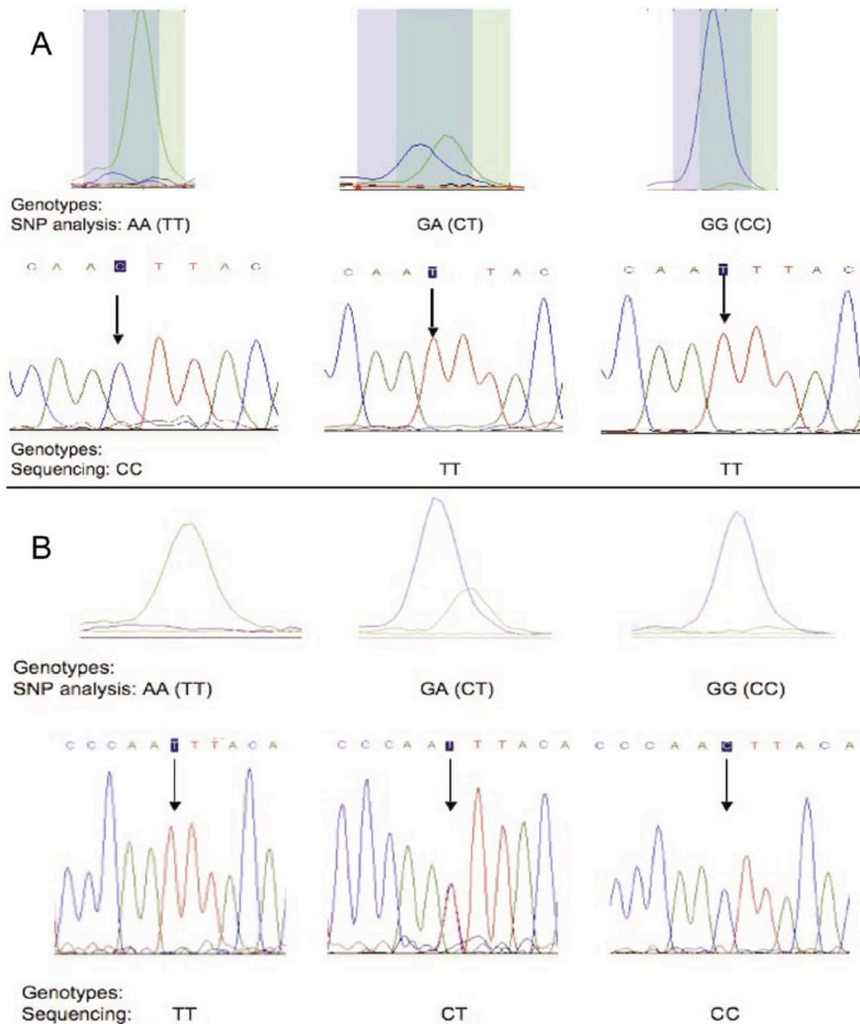

Figure 1. The SNP rs3738579 genotype of selected DNA samples established by the SNaPshot method and by direct sequencing. (A) Examples demonstrating the discrepancies between the SNP and the sequencing analysis. The green and blue background is software-derived and denotes the expected position of the two alleles A and G. Note that the SNP genotype is assessed from the reverse DNA strand. (B) Examples demonstrating agreement between the SNP rs3738579 assessments and sequencing analysis.

Table 1. Primer Sequences for Amplification of the Genomic Fragments and Assessment of the SNP Genotypes

SNP	Polymorphism	Primers	Primer sequences 5′-3′
rs10774671	G/A	SNP	CATGTGTCTCACCCTTTCA
		*OAS1*f	ATGGCATGTCACAGTGTC
		*OAS1*r	CATAGAAGGCCAGGAGTC
rs1051042	C/G	SNP	AGACCGACGATCCCAGGA
		*OAS1*f	ATGGCATGTCACAGTGTC
		*OAS1*r	CATAGAAGGCCAGGAGTC
rs2660	G/A	SNP	GGGACTCTTGATCCAGAGA
		*OAS1*f	ATGGCATGTCACAGTGTC
		*OAS1*r	CATAGAAGGCCAGGAGTC
rs3738579	C/T	SNP	TCAAAGAAGCTTTGAGTGTAA
		*RNASEL2*r	CTCATTGACATCTGCTC
		*RNASEL2*f	GATTCAAGTGTTTTCTCCC

SNP primers were separated by capillary electrophoresis using the ABI 3130×l Genetic Analyzer and the polymer POP-7 (both from Applied Biosystems). The results were analyzed using the GeneMapper software.

Initially, the quality of each specific SNP assay was evaluated by processing 24–32 different control samples through the SnaPshot protocol. The criteria for a successful genotype assessment comprised correct variable nucleotides, correct length of product, no background or unspecific peaks, and proper height and shape of the peaks. If the peaks are low or if the sample is overloaded, the result can be misinterpreted. If one or more of these criteria were not met, a design of a new SNP primer from the opposite DNA strand may solve the problem. When providing a satisfactory result, the SNP primers were pooled adding four to seven in each reaction and processing through the SNaPshot protocol.

For SNPs with a low mutation rate only the heterozygous samples were sequenced and no discrepancy was found. A small fraction of samples (20–25) of the polymorphic SNPs was sequenced to verify each of the genotypes. To our surprise we found that a subset of the SNP rs3738579 genotypes, scored to be clearly homozygous for cytosine, were either heterozygous (C/T) or homozygous for thymine. Furthermore, some of the heterozygous samples were homozygous for thymine. Only a trace of the second allele for the homozygous/heterozygous samples was seen in the SNaPshot analysis (Figure 1). The observation that certain SNPs, presenting an unambiguous result after the SNaPshot analysis, had a different genotype when sequenced was highly unexpected and led to subsequent re-analysis of 258 random samples to evaluate the genotype assessment of SNP rs3738579. In total, 16 samples (6.2%) presented a different genotype (Table 2).

A similar problem was encountered for SNPs: rs10774671, rs1051042, and rs2660 within *OAS1*. All presented a clear SNP result, no

background noise, and peaks of satisfactory size, and yet a subset had a genotype different from the genotype found by the sequencing analysis (Figure 2). No discrepancies were found for assessment of the remaining 12 SNP genotypes.

To ensure a uniform template for subsequent analysis, a subset of DNA fragments containing each of the four SNPs (rs3738579, rs10774671, rs1051042, and rs2660) were PCR-amplified and divided in two; one for sequencing and one for SNP analysis. Of 16 randomly chosen samples reanalyzed for the 3 *OAS1* SNP genotypes, 15 were reproducible but one sample provided contradictive genotypes. In total the rate of misinterpretation of the 4 SNPs was between 6% and 15% (Table 2).

Initial focus was directed to the SNP primer design. We found no evidence supporting the possibility that the primer binds to other sequences in the amplified fragment or in the genome (Figure 3). Neither did we find evidence that loops would facilitate annealing of the primer to a proximate sequence, thereby being extended with a different nucleotide.

For each case the resulting chromatogram provided the correct size of the extended primer and the expected two variable nucleotides, strongly indicating that the primer was full-length and annealed to the correct target sequence.

If the SNP results are contradictive or non-assessable, the primer can be synthesized complementarily to the opposite DNA strand, which solves the problem in most cases. For rs3738579, we found a new mutation 17 bp upstream from this SNP in patients diagnosed with uterine cervical cancer. This mutation was not reported to the databases (Human Genome Browser and NCBI; genome.ucsc.edu), so the frequency in the Danish population was unclear. We therefore designed the SNP primer from the opposite DNA strand. A successful result was obtained after redesigning the SNP primer for rs3177979 from the forward strand as the initial assay provided three alleles.

DNA that is of degraded and of poor quality could provide an explanation of the genotyping ambiguity. Tumor DNA can be heterogeneous, one allele deleted or amplified, or microdeletions may block for allelic amplification. We used leukocyte DNA, gently purified to diminish double-stranded DNA breaks, and the alleles were likely to be present in an equal number, the tissue being non-malignant.

We and others used a reduced amount of the SNaPshot mix (3 µL instead of the recommended 5 µL). Thirty-two samples were reanalyzed using the recommended amount, and the result was unambiguous (data not shown).

The results presented here state how important it is to sequence a fragment containing each SNP for a subset of the samples, even if the results seem unambiguous at first and give no reason to doubt the genotype assessment. We have no reason to believe that the primer extension methodology is failing, but we are more of the opinion that locus-

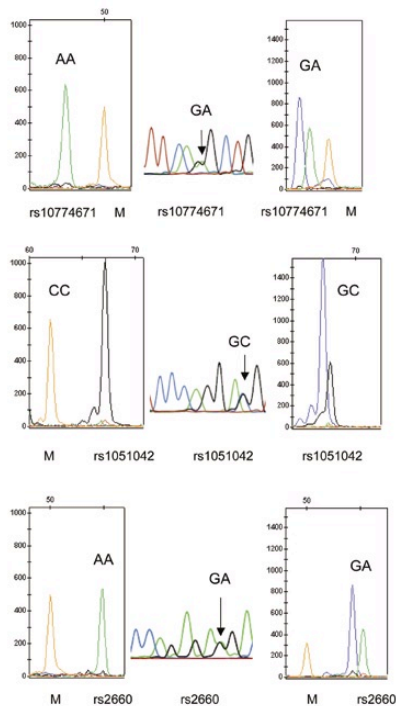

Figure 2. Examples of disagreement between different SNP analysis of the same sample and sequencing results for SNPs: rs10774671, rs1051042, and rs2660. The orange peak is a size marker.

Table 2. Reanalysis of 67 Samples for the *OAS1* SNPs: rs10774671, rs2660 and rs2660, and 258 Samples of the RNASEL SNP rs3738579

Genotype change	rs10774671	rs2660 no. (no./67 sequenced)	Genotype change	rs2660	Genotype change	rs3738579
GA–GG	3 (4.5%)	–	CG–GG	1 (1.5%)	CC–TC	5 (2.0%)
GA–AA	6 (9.0%)	Four (6.0%)	CG–CC	3 (4.5%)	TC–TT	9 (3.5%)
GG–GA	1 (1.5%)	–	GG–GC	2 (3.0%)	TT–CC	(0.4%)
					CC–TT	(0.4%)
In total	15.0%	6.0%		8.9%		6.2%

*No. refers to the number of samples.

OAS1

gcaggaagactccctgatgtgatcatgtgtctcaccctttcaggctgaaa rs10774671
gcaacagtgcagacgatgagaccgacgatcccaggaggtatcagaaatat rs1051042
ggttacattggaacacatgagtaccctcatttctctcatagacccagcac
actccaggcagcatccaccccacaggcagaagaggactggacctgcacca
tcctctgaatgccagtgcatcttgggggaaagggctccagtgttatctgg
accagttccttcattttcaggtgggactcttgatccagagaggacaaagc rs2660
tcctcagtgagctggtgtataatccaggacagaacccaggtctcctgact

RNASEL

ctgctgctctgttgccagagaatcccaatttacactcaaagcttctttga rs3738579
ttaagtgctaggagataaattttgc

Figure 3. The location of SNP primers within OAS1 exon 7 and 5′ UTR of RNASEL. The polymorphic bases are in red, and arrows illustrate the orientation and extension of each primer.

specific copy number variation or structures in the DNA may influence the obtained results.

ACKNOWLEDGMENTS

T.E. Keldsen is acknowledged for excellent technical assistance. The Danish Cancer Research Foundation and The Aarhus University Research Foundation supported this work. The Danish Cancer Society supported C.W.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1. **Altshuler, D., L.D. Brooks, A. Chakravarti, F.S. Collins, M.J. Daly, and P. Donnelly.** 2005. A haplotype map of the human genome. Nature *437*:1299-1320.
2. **Rocha, D., I. Gut, A.J. Jeffreys, P.Y. Kwok, A.J. Brookes, and S.J. Chanock.** 2006. Seventh international meeting on single nucleotide polymorphism and complex genome analysis: 'ever bigger scans and an increasingly variable genome.' Hum. Genet. *119*:451-456.
3. **Kim, S. and A. Misra.** 2007. SNP genotyping: technologies and biomedical applications. Annu. Rev. Biomed. Eng. *9*:289-320.
4. **Hansen, L.L., M. Yilmaz, J. Overgaard, J. Andersen, and T.A. Kruse.** 1998. Allelic loss of 16q23.2-24.2 is an independent marker of good prognosis in primary breast cancer. Cancer Res. *58*:2166-2169.
5. **Hansen, L.L., J. Andersen, J. Overgaard, and T.A. Kruse.** 1998. Molecular genetic analysis of easily accessible breast tumour DNA, purified from the leftover from hormone receptor measurement. APMIS *106*:371-377.

Received 13 June 2007; accepted 12 October 2007.

Address correspondence to Lise Lotte Hansen, Institute of Human Genetics, The Bartholin Building, University of Aarhus, DK-8000 Aarhus C, Denmark. e-mail: lotte@humgen.au.dk

To purchase reprints of this article, contact: Reprints@BioTechniques.com