European Journal of Human Genetics (2000) 8, 933–938

www.nature.com/ejhg

ARTICLE

Determination of the frequencies of ten allelic variants of the Wilson disease gene (*ATP7B*), in pooled DNA samples

Charlotta Olsson[1,2], Erik Waldenström[1], Kerstin Westermark[1], Ulf Landegren[1,2] and Ann-Christine Syvänen[1]

Molecular Medicine, [1]Department of Medical Sciences, [2]Department of Genetics and Pathology, Uppsala University, Sweden

Wilson disease is an autosomal recessive disorder characterised by toxic accumulation of copper in liver, brain and other organs. The disorder is caused by mutations in the *ATP7B* gene, encoding a copper transporting P-type ATPase. Based on the number of known patients with this diagnosis in Sweden, the prevalence can be estimated to 1 in 250 000 to 300 000, whereas the prevalence of Wilson disease has been estimated to be 1 in 30 000 in other populations. We estimated the prevalence of Wilson disease by determining the Swedish population frequencies of two mutant alleles, making up approximately half the mutations in Swedish Wilson patients, in a large number of DNA samples. In addition we determined the allele frequencies of eight common single-nucleotide polymorphisms (SNPs) in the *ATP7B* gene. For the analyses we devised two strategies for analysing pooled DNA samples using the quantitative minisequencing method. The two procedures allowed sensitive identification of rare mutant alleles present as a mixture with an excess of the normal allele, as well as accurate estimation of the frequencies of the common SNPs in a large pooled DNA sample. *European Journal of Human Genetics* (2000) **8**, 933–938.

Keywords: Wilson disease; *ATP7B* gene; minisequencing; DNA pools; single-nucleotide polymorphisms

Introduction

Wilson disease is an autosomal recessive disorder characterised by failure to incorporate copper into ceruloplasmin in the liver and to excrete copper from the liver to the bile, resulting in toxic copper accumulation predominantly in liver and brain.[1,2] This results in acute and chronic liver disease and/or neurologic dysfunction. Diagnosis of Wilson disease is based on lowered levels of ceruloplasmin and increased free copper levels in urine and serum as well as on elevated copper concentrations in the liver. Wilson disease is lethal if left untreated, but if diagnosed, progression of the disease can be interrupted. However, as the typical biochemical findings may also be observed in other hepatic disorders, Wilson disease can sometimes be difficult to diagnose. Furthermore, some patients neither show clinical signs of neurologic dysfunction nor do they have the typical corneal deposits of copper, the so-called Kayser-Fleischer rings.[3]

The gene causing Wilson disease was cloned in 1993[4–6] and named *ATP7B*. The protein encoded by this gene is a copper transporting P-type ATPase, which is predominantly expressed in liver and brain.[7] Almost 200 mutations in the *ATP7B* gene have been reported world wide (http://www.medgen.med.ualberta.ca/database.html). The two most common mutations in Swedish patients are a C–A transversion in exon 14 (C3207A) and a C–T transition in exon 13 (C2930T), accounting for 33% and 11% of the Swedish disease alleles, respectively.[8] The prevalence of Wilson disease in Sweden has been estimated to be 1 in 250 000, whereas its prevalence has been estimated to be 1 in 30 000 in other populations.[9] In the present study we determined the frequency of the two major disease-causing alleles in the *ATP7B* gene in a large Swedish population sample to investigate if the 10-fold lower prevalence of Wilson disease

Correspondence: Ann-Christine Syvänen PhD, Molecular Medicine, Department of Medical Sciences, Entrance 70, 3rd Floor, Research Department 2, Uppsala University Hospital, SE-751 85 Uppsala, Sweden. Tel: +46 18 66 29 59; Fax: +46 18 66 25 19; E-mail: ann-christine.syvanen@medsci.uu.se
Received 31 May 2000; revised 19 July 2000; accepted 31 July 2000

could be due to patients being missed in the clinic. In order to simplify detection of rare disease alleles we used a strategy in which pools of 10 DNA samples were analysed for the presence of the two major disease-causing mutations in Sweden. The *ATP7B* gene also contains several single-nucleotide polymorphisms (SNPs) that are more common in the population than the disease-causing mutations.[8,10] To determine the Swedish allele frequencies of eight of the common SNPs in coding regions of the *ATP7B* gene in the same sample material, we devised a different pooling strategy. A large pool, consisting of 2500 Swedish DNA samples, was analysed quantitatively to determine the proportion of the two sequence variants of each SNP, which reflects directly the allele frequencies at these sites in the Swedish population. In both strategies the pooled DNA samples were analysed by the solid phase minisequencing method.[11] This PCR-based primer extension method identifies single nucleotide variations by incorporation of a single labelled nucleotide analogue. The analysis showed that the minisequencing method allows sensitive detection of minority sequence variants and accurate quantitative analysis of sequence variants present as a mixture in a pooled DNA sample.

Material and methods

DNA samples

DNA was extracted from 2640 anonymous Swedish blood samples. DNA was extracted from the blood samples using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA). The DNA concentration of the samples was determined by measuring the absorbance at 260 nm in a spectrophotometer (DU 530 Beckman Instruments Inc, Fullerton, CA, USA). First, equal amounts of DNA from 10 individuals were combined into pools with DNA concentrations of 200–400 ng/μl. These pools of 10 samples were further combined into a large 'super pool' containing an equal amount of DNA from 2500 individuals. The rare alleles were detected in pools of 10 samples and the frequencies of common SNP alleles were determined from the 'super pool'.

PCR amplification

The PCR primers were designed using the OLIGO program (Molecular Biology Insights Inc, Cascade, CO, USA) based on unpublished intronic sequences of the *ATP7B* gene, kindly provided by Drs Conrad Gilliam and Konstantin Petrukhin (Columbia University, New York). The sequence of the 5'-biotinylated forward primer for exon 13 was 5'-ACT CTG TCC TGT TTT CAG AAC CC-3' and that of the reverse primer was 5'-CGT GCT ACA GGC TGA CCT TGT GC-3'. The forward 5'-biotinylated primer for exon 14 was 5'-AGA CTG TGA TGT TTG ACA AGA CTG G-3' and the reverse primer was 5'-CCA AGT CCA CGT ACC TCT TTA-3'. The sequences of the primers for amplifying the exons containing eight SNPs in the *ATP7B* gene are given in Waldenström *et al*.[8] The DNA fragments containing the disease mutations or SNPs in the *ATP7B* gene were amplified from 200–400 ng of DNA using 1.25 units of *Taq* DNA polymerase, 200 μM dNTPs, the biotinylated primer at 0.2 μM concentration and the other primer at 1 μM concentration in 100 μl of DNA polymerase buffer (50 mM KCl, 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$, 12.5 mg/ml bovine serum albumin, PE Biosystems, Foster City, CA, USA). The temperature was varied during 35 cycles of 94°C for 30 s, an annealing temperature between 55°C and 65°C depending on the T_m of the primers, for 30 s and 72°C for 45 s.

Minisequencing reaction

Four 10 μl aliquots of each PCR product were mixed with 40 μl of binding buffer (50 mM phosphate buffer pH 7.5 containing 0.15 mM NaCl and 0.1% Tween 20) in streptavidin-coated microtiter plate wells (Combiplate 8, Labsystems, Helsinki, Finland) and incubated at 37°C for 1.5 h in a shaker (Thermomix 1415, Labsystems). The wells were washed 6 times with 40 mM Tris-HCl, pH 8.8, 1 mM EDTA, 50 mM NaCl and 0.1% Tween 20 in a plate washer (Wellwash, Labsystems). The non-biotinylated strand of the PCR-product was removed by denaturation with 60 μl of 0.1 M NaOH for 3 min. After washing as above, 50 μl of a minisequencing reaction mix containing DNA polymerase buffer, 0.05 U of *Taq* polymerase (PE Biosystems), 0.1 μCi of the appropriate [^{3}H]-dNTP ([^{3}H]-dATP, TRK 633, [^{3}H]-dCTP, TRK 625, [^{3}H]-dGTP, TRK 627, [^{3}H]-dTTP, TRK 576, Amersham Pharmacia Biotech, Amersham, UK) and 10 pmol of the appropriate detection primer was added to the wells (Table 1). The plates were incubated at 50°C for 10 min. The unincorporated label was removed by washing as above, and the detection primers were released with 100 μl of 0.1 M NaOH, and measured in a liquid scintillation counter (1414, Wallac, Turku, Finland).

Table 1 Position of the variant nucleotides and the minisequencing primers on the *ATP7B* gene

Positions of mutations and polymorphisms (nt number)[a]	*Exon*	*Primer position (nt number)*[a]
C2930T[b]	13	2953–2931[c]
C3207A[b]	14	3230–3208[c]
G1216T	2	1196–1215
C1366G	3	1346–1365
G2495A	10	2475–2494
G2855A	12	2835–2854
G2973A	13	2953–2972
G3009A	13	2989–3008
G3045A	13	3025–3044
C3419T	16	3399–3418

[a]The nucleotide numbering is according to the published ATP7B cDNA sequence (accession no. U03464, *GenBank*); [b]the two WD causing mutations analysed in 2640 individuals; [c]the non-coding strand was analysed.

Interpretation of the result

The result of the minisequencing assay is expressed as the ratio between the signals from the incorporated [^{3}H]-dNTP corresponding to the two sequence variants at each analysed site. A DNA sample with a 1:20 ratio between the mutant and normal sequence (5%) was included in each series of pools as a reference sample for identification of DNA pools containing one mutant allele among the 20 alleles in each pool. The presence of a rare mutant allele in a pooled DNA sample is revealed by a signal ratio that is in the same range as that of the 5% reference sample, and elevated by more than two standard deviations compared with the ratio in a homozygous normal sample with 0% of the mutant allele. The allele frequencies (f) of the common SNPs were calculated by comparing the signal ratios observed in the large pooled sample (R_{pool}) to the corresponding signal ratios in heterozygous samples (R_{Het}), where the two alleles are present at a 1:1 ratio (50%) according to the formulae:

$$f_{allele1} = \frac{R_{pool}/R_{Het}}{1+R_{pool}/R_{Het}};\ f_{allele2} = \frac{1}{1+R_{pool}/R_{Het}}$$

Results

In the present study we determined the Swedish population frequencies of the two major mutant alleles of the *ATP7B* gene which cause the recessively inherited Wilson disease, and the population frequencies of eight common single polymorphisms (SNPs) in the *ATP7B* gene. The allele frequencies of both the rare alleles due to mutations in exon 13 and 14 of the *ATP7B* gene and those of the eight more common SNPs were determined by analysing pooled DNA samples quantitatively by competitive PCR using the solid-phase minisequencing method. Two different pooling strategies were employed.

Detecting rare alleles

To estimate the optimal pool size for detecting rare Wilson disease-causing alleles, two standard curves were constructed by analysing samples containing mixtures of known amounts of the mutant and normal sequence in exons 13 and 14 (Figure 1). As can be seen from the standard curve in Figure 1A, samples containing significantly less than 1.5% of the sequence with a mutation in exon 13 can be distinguished from a homozygous normal sample (with 0% of the mutant sequence). The sensitivity of detecting the mutation in exon 14 of the *ATP7B* gene was about 3% (Figure 1B). For both mutations there is a linear relationship between the ratio of the two sequence variants in the sample and that of the labelled allele-specific nucleotides incorporated in the minisequencing reactions from the detection limit up to at least 50% of the mutant sequence. Based on the observed sensitivity, equal amounts of DNA were pooled from sets of 10 samples to allow for errors in the DNA quantification,

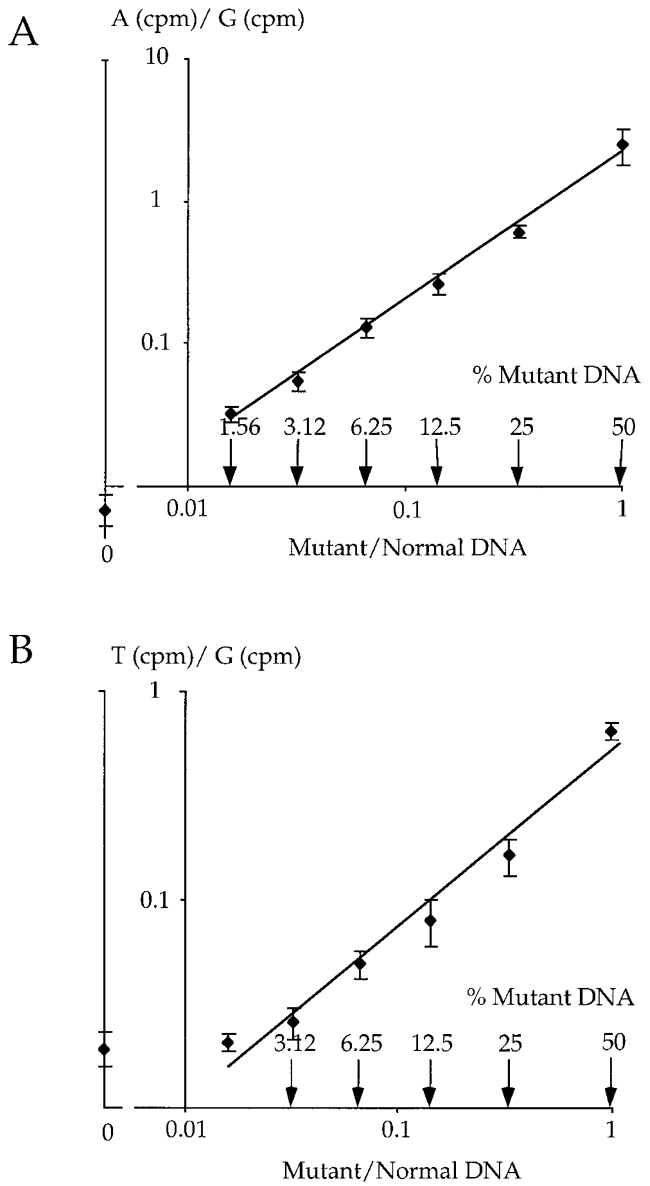

Figure 1 Solid-phase minisequencing standard curves. Minisequencing signal ratio plotted as a function of the ratio between the sequence containing **(A)** the C2930T mutation in exon 13 and **(B)** the C3207A mutation in exon 14 of the *ATP7B* gene and the normal sequence in mixed samples. Mean value and standard deviations of four parallel assays are shown. Arrows indicate the corresponding % of mutant sequence in the mixtures. The signal ratios and their standard deviations in control samples without mutant sequence are given on the line to the left of the y axis. The nucleotide corresponding to the non-coding DNA-strand was incorporated in the minisequencing reactions.

although the sensitivity determined from the standard curves would have allowed larger pools. With this pool size the mutant sequence would constitute 5% in pools including a single carrier individual. Altogether 264 pools were analysed

for the two rare mutations in the *ATP7B* gene. The analysis revealed clearly elevated signal ratios compared with those in pools without a mutant allele and corresponding to 5% of DNA with a mutation in exon 14 of the *ATP7B* gene in four of the pools (Pool J, M, 9 and 192 in Table 2). An elevated signal ratio in pool 158 (Table 2) indicated that it contained one heterozygous sample with the exon 13 mutation of the *ATP7B* gene. The individual DNA samples in pools showing elevated signal ratios were analysed individually, and a heterozygous carrier of the Wilson disease mutation was identified in each of them. Based on this analysis, the Swedish population frequency of the mutant allele in exon 14 is expected to be 1/1300 with a 95% confidence interval between 1/560 and 1/4400. The frequency of the exon 13 mutation is 1/5300 with a wide 95% confidence interval of between 1/1200 and 1/100 000. The frequencies of the exon 14 and 13 mutations estimated from the number of patients and the proportion of these mutations among the patients in Sweden are 1/760 and 1/2300, respectively.

Detecting common alleles or SNPs

The coding sequence of the *ATP7B* gene contains eight common SNPs of which three are silent, and five are missense changes.[8] The frequencies of these sequence variants were determined rapidly and accurately using a different pooling strategy. An equal amount of DNA from 2500 individuals out of the 2640 analysed was combined into a 'super pool', which was subjected to quantitative analysis by the minisequencing method for each of the eight SNP sites. The ratios of the sequence variants in the super pool, corresponding to their population frequencies, were estimated from the ratio between the ^{3}H-labeled dNTPs incorporated in the minisequencing reactions. The minisequencing ratio is influenced if the sequence immediately downstream of the analysed nucleotide position contains more than one of the same nucleotide, so that more than one ^{3}H-labeled dNTP is incorporated for this allelic variant. That is the case for the G2495A and C3419T polymorphisms (Table 3). In practice, the ratio is also affected by any sequence dependent misincorporation of ^{3}H-labeled dNTPs by the DNA polymerase that may occur. To correct for these sequence context-dependent factors, heterozygous samples, in which the two alleles are known to be present in equal amounts, were used as quantification standards for each of the analysed SNPs (Table 3). The allele frequencies of the SNPs were also determined by genotyping 20 individual DNA samples at each SNP. This analysis gave a result concordant with the quantitative analysis of the 'super pool'. Table 3 shows the Swedish allele frequencies of the eight SNPs.

Table 2 Result from the minisequencing analysis of pools containing DNA from Wilson disease carriers compared with negative pools and reference samples

	Signals (cpm)[a]		Signal ratio[b] Mutant/Normal		
	Mutant	Normal	Pos pool	Neg pool	5% reference
Exon 13 G2930A					
Pool 158	215±26	1340±77	0.16	0.010	0.23
Exon 14 G3207T					
Pool J	420±21	4400±530	0.096	0.0079	0.059
Pool M	160±3	5270±340	0.031	0.0079	0.059
Pool 9	120±40	2670±73	0.045	0.0067	0.042
Pool 192	190±48	2280±8	0.082	0.018	0.065

[a]Mean value and standard deviations of four parallel minisequencing assays of the same PCR product; [b]mean signal ratio in pools that were interpreted as negative for the analysed mutation and a reference sample with 5% of the mutant sequence in the same experiment.

Discussion

We applied two different pooling strategies for rapid and accurate determination of the Swedish population frequencies of ten sequence variants in the *ATP7B* gene. The frequencies of the two most common mutant alleles of the *ATP7B* gene causing Wilson disease in Sweden[8] were determined by analysis of pooled samples, each of which contained an equal amount of DNA from 10 individuals. The number of identified carriers of these two *ATP7B* gene mutations coincides with the expected carrier frequency based on the number of diagnosed patients with Wilson disease and on the distribution of the exon 13 and 14 mutations among these patients.[8] Our study supports the view that Wilson disease is well diagnosed in Sweden, and that the disease is really more rare in Sweden than in other countries, although the number of identified carriers was small.

A pooling approach has been proposed previously for population screening of the cystic fibrosis three base pair deletion mutation ΔF508 by PCR-based heteroduplex analysis. A model for calculating the most cost-effective number of samples to be pooled based on the population frequency of the mutation to be detected was also presented.[12] For rare mutant alleles, such as the Wilson disease alleles, the optimal number of samples per pool is in practice determined by the sensitivity of the method applied for detecting the minority allelic variant in an excess of the normal sequence. For hybridisation and sequencing methods the detection limit for genomic sequence variants is 10% or higher.[13–15] The minisequencing method has proved to be more sensitive, as illustrated in a previous study, in which we were able to detect a rare recessive disease allele at the 1% level in pooled DNA samples.[16] In detecting the two Wilson disease alleles, the detection limit of the minisequencing method was lower than 1% to 3%, respectively. Based on our experience with

Table 3 Allele frequencies of SNPs in the *ATP7B* gene, in a DNA pool from 2500 individuals and in 20 random individuals

	Signals (cpm)[a]			Allele distribution Normal–Variant	
	Normal	Variant	Signal ratio Normal/Variant	Super pool	Individual samples
G1216T					
Super pool	1120±330	1940±6	0.58	0.44–0.56	0.46–0.54
Heterozygote	1520±100	2020±150	0.75		
C1366G					
Super pool	1800±170	1180±210	1.52	0.52–0.48	0.50–0.50
Heterozygote	1870±4	1340±30	1.40		
G2495A					
Super pool	2390±540	160±50	14.9	0.49–0.51	0.46–0.54
Heterozygote	2780±50	180±10	15.4		
G2855A					
Super pool	3500±220	9460±250	0.37	0.47–0.53	0.33–0.66
Heterozygote	3760±250	9150±560	0.41		
G2973A					
Super pool	840±60	170±30	4.94	0.92–0.08 0.94–0.06[b]	0.90–0.10
Heterozygote	820±60	1820±70	0.45		
5% A-allele	1390±180	250±20	5.56		
G3009A					
Super pool	200±20	30±5	6.67	0.87–0.13 0.89–0.11[b]	0.85–0.15
Heterozygote	260±90	240±20	1.08		
5% A-allele	340±20	20±2	17		
G3045A					
Super pool	2000±160	80±8	25	0.96–0.04 0.97–0.03[b]	0.90–0.10
Heterozygote	1630±130	1600±180	1.01		
5% A-allele	1970±310	150±10	13.1		
C3419T					
Super pool	20800±1070	3300±180	6.30	0.53–0.47	0.74–0.26
Heterozygote	19700±1700	3490±140	5.64		

The nucleotides denoted as 'normal' are those given in the published sequence of the *ATP7B* gene (accession no. U03464, *GenBank*); the normal nucleotide is given on the left, followed by the nucleotide number in the *ATP7B* sequence; [a]mean value and standard deviations of four parallel assays from the same PCR product; [b]calculated using a 5% allele mixture as reference.

other mutations the detection limit is in most cases close to 1% when ^{3}H-dNTPs are used as label, but higher with other labelled nucleotide analogues.[17]

Currently there is much effort to discover dense sets of SNPs to serve as markers to identify genes underlying common, multifactorial disorders by genome-wide linkage disequilibrium mapping or analysis of candidate genes. In the year 2000 SNP consortium has catalogued 300 000 SNPs.[18] Once the SNPs have been discovered, population-specific allele frequency databases will need to be established for them. To create allele frequency databases and perform association studies, a significant increase in throughput and reduction in costs of the current available assays for typing SNPs is needed. One means would be to analyse alleles quantitatively in pooled DNA samples containing DNA from a large number of case and control individuals instead of genotyping individual DNA samples.

The concept of DNA sample pooling to reduce labour and costs in an association study was first suggested in 1985 by Arnheim *et al*.[19] In that study Southern blot hybridisation was used to search for restriction fragment length polymorphisms in the HLA class II locus that were in linkage disequilibrium with susceptibility to insulin-dependent diabetes in pooled samples containing DNA from tens of individuals. In several subsequent studies allele frequencies of microsatellite markers were determined by analysis of pooled DNA samples using PCR, followed by size separation and quantification of alleles.[20–22] We earlier determined quantitatively the allele frequencies of a panel of SNPs to be used in forensics in large pooled samples containing DNA from several hundred individuals by competitive PCR using solid-phase minisequencing.[11] In the present study the method proved accurate for quantification of eight common variant alleles in the *ATP7B* gene in a large 'super pool' of 2500 individuals.

We conclude that minisequencing primer extension is a suitable reaction principle for future cost-effective population-based screening of rare mutations, for determination of

population frequencies of SNPs and for association studies using pooled samples on a large scale.

Acknowledgements

We thank Drs Joe Terwilliger, Anders Isaksson, Maria Lagerström and Sara Wedrén for their helpful contributions. The study was supported by grants from the Swedish Research Council for Medical Sciences.

References

1 Bull PC, Cox DW: Wilson disease and Menkes disease: new handles on heavy-metal transport. *Trends Genet* 1994; **10**: 246–252.
2 Danks DM: *The Metabolic Basis of Inherited Disease.* McGraw-Hill: New York, 1989, pp 1411–1431.
3 Gollan JL, Gollan TJ: Wilson disease in 1998: genetic, diagnostic and therapeutic aspects. *J Hepatol* 1998; **28**: 28–36.
4 Bull PC, Cox DW: Long range restriction mapping of 13q14.3 focused on the Wilson disease region. *Genomics* 1993; **16**: 593–598.
5 Tanzi RE, Petrukhin K, Chernov I *et al*: The Wilson disease gene is a copper transporting ATPase with homology to the Menkes disease gene. *Nat Genet* 1993; **5**: 344–350.
6 Yamaguchi Y, Heiny ME, Gitlin JD: Isolation and characterization of a human liver cDNA as a candidate gene for Wilson disease. *Biochem Biophys Res Commun* 1993; **197**: 271–277.
7 Bull PC, Thomas GR, Rommens JM, Forbes JR, Cox DW: The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. *Nat Genet* 1993; **5**: 327–337.
8 Waldenström E, Lagerkvist A, Dahlman T, Westermark K, Landegren U: Efficient detection of mutations in Wilson disease by manifold sequencing. *Genomics* 1996; **37**: 303–309.
9 Terada K, Schilsky ML, Miura N, Sugiyama T: ATP7B (WND) protein. *Int J Biochem Cell Biol* 1998; **30**: 1063–1067.
10 Thomas GR, Forbes JR, Roberts EA, Walshe JM, Cox DW: The Wilson disease gene: spectrum of mutations and their consequences. *Nat Genet* 1995; **9**: 210–217.
11 Syvänen A-C, Sajantila A, Lukka M: Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. *Am J Hum Genet* 1993; **52**: 46–59.
12 Gille C, Grade K, Coutelle C: A pooling strategy for heterozygote screening of the delta F508 cystic fibrosis mutation. *Hum Genet* 1991; **86**: 289–291.
13 Farr CJ, Saiki RK, Erlich HA, McCormick F, Marshall CJ: Analysis of RAS gene mutations in acute myeloid leukemia by polymerase chain reaction and oligonucleotide probes. *Proc Natl Acad Sci USA* 1988; **85**: 1629–1633.
14 Lai E, Riley J, Purvis I, Roses A: A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. *Genomics* 1998; **54**: 31–38.
15 Ahmadian A, Lundeberg J, Nyrén P, Uhlen M, Ronaghi M: Analysis of the p53 tumor suppressor gene by pyrosequencing. *Biotechniques* 2000; **28**: 140–144, 146–147.
16 Syvänen A-C, Ikonen E, Manninen T *et al*: Convenient and quantitative determination of the frequency of a mutant allele using solid-phase minisequencing: application to aspartylglucosaminuria in Finland. *Genomics* 1992; **12**: 590–595.
17 Syvänen A-C: From gels to chips: 'minisequencing' primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum Mutat* 1999; **13**: 1–10.
18 Masood E: A consortium plans free SNP map of human genome [news]. *Nature* 1999; **398**: 545–546.
19 Arnheim N, Strange C, Erlich H: Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of the HLA class II loci. *Proc Natl Acad Sci USA* 1985; **82**: 6970–6974.
20 Lipkin E, Mosig MO, Darvasi A *et al*: Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage. *Genetics* 1998; **149**: 1557–1567.
21 Pacek P, Sajantila A, Syvänen A-C: Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples. *PCR Methods Appl* 1993; **2**: 313–317.
22 Shaw SH, Carrasquillo MM, Kashuk C, Puffenberger EG, Chakravarti A: Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. *Genome Res* 1998; **8**: 111–123.