Challenges of genetic diversity and the use of high throughput genotyping in genetic epidemiology

Vessela N. Kristensen

Department of Genetics, Institute for Cancer Research, the Norwegian Radium Hospital, Montebello, 0310 Oslo
Telephone: *47 22 93 44 17(21)   Telefax: *47 22 93 44 40   E-mail: vessela@ulrik.uio.no   http://www.med.uio.no/dnr/Inst/genetics/genetics.htm

ABSTRACT

This paper provides a discussion of the challenges in studies of genetic variation in epidemiology represented by the vast complexity of genetic variation. Choices of design and targets of the right genetic markers could be essential for detection of genetic susceptibility. Identification of genetic variation that is likely to be involved in the biological pathway of a particular condition could be of vital importance. A review is given of current use of genetic variation of a series of genes *(TP53, GST, CYP)* in studies of genetic susceptibility. Approaches to studies of gene-environment interaction involving exposure to endocrine disrupters are discussed. Finally the paper discusses new developments in the methodology for high throughput genotyping. Such techniques are assumed to be of particular relevance for larger epidemiological studies.

INTRODUCTION

Large-scale genotyping may enable identification of single nucleotide polymorphisms (SNPs) which may be relevant for further studies in relation to susceptibility markers, clinical and histopathological characteristics of disease, prognosis and response to treatment. To be able to achieve these goals we need to evaluate and introduce methods for high throughput SNP analysis. We hope to be able to establish genetically predisposed variations in biologically relevant pathways involved in given environmental and life-style characteristics, exposure patterns, clinical and histopathological phenotypes as well as response to treatment.

CONCEPTUAL PROBLEMS OF STUDYING AND REDUCING GENETIC VARIATION

Molecular epidemiology faces the challenge to explain complex phenotypes resulting from the interaction of complex genotypes with complex environments, thus preceding the similar goals of functional genomics. One of the greatest challenges is the vast *complexity of genetic variation*. SNPs occur with a frequency of 1/1250 bp and in a total number of 3.12 billion nucleotides they may be estimated to be about 2.5 million. The most obvious candidates to influence the function, the coding SNPs (cSNPs), roughly estimated to 100 000 will, allowed for two different nucleotides at a position, result in 3^{100000} possible genotype combinations for every individual, again giving room for the biblical number of around 10^{50000} genetically different individuals in respect to these SNPs (1). This illustrates the problems of studying genetic variation in epidemiological studies.

Epidemiological studies therefore have to use different strategies to reduce the complexity in search for particular relevant SNPs. The main alternative strategies are briefly summarized in Table 1.

The "biological" candidate gene approach searches for SNPs in genes with key functions for metabolism of xenobiotics (i.e. cytochrome P450, Glutathione S transferases, N-acetyltransferases, etc.), which may be responsible for altered proximal phenotypes (mRNA and protein levels and activities). Alternatively, the "positional" approach suggests to define the hot spots of recombination and to reduce the complexity by defining all SNPs in linkage disequilibrium in between (LD domains) as the next smallest building brick of the variation skeleton (2). Whether they can always be well defined, stable in different populations and of a manageable size is still an open question. To reduce the populational complexities one could look back to classical family studies and let the family trees create

Table 1. Alternative strategies for reducing the complexity of the search for molecular markers.

Whole genome global analysis	Hypothesis driven "reductionistic" approach
Positional approach, based on chromosomal location	Candidate gene approach (based on biological function)
"common" alleles associated with disease	"rare" alleles associated with disease
"family" studies	"case-control" populations
SNP sets according to their location in LD blocks	SNP sets according to their location in coding sequences

the initial pattern from the chaos as theoretically predicted (3) and experimentally proved on the example of Crohn's disease (4). When looking for susceptibility markers to disease, both common and rare alleles have been considered. Given the lack of equilibrium in the human population, due to the rapid expansion from a small and isolated pool, it may be argued that even common alleles may be useful susceptibility markers (5). According to other population genetics studies, rare alleles have a better chance to make a difference if deleterious, their low frequency in itself incriminating them for possible harmful phenotype (6).

In addition to the problems of tackling *genetic heterogeneity*, epidemiological studies also need to tackle *heterogeneity of complex phenotypes* such as diabetes, cardiovascular diseases and cancer. Identification of relevant well-defined phenotypic groups is at least as complex as identification of the correct set of SNPs as susceptibility markers, and without relatively homogeneous groups of cases, chances of detecting association with genetic variation may fade. In oncological diseases, the phenotypes could with decreasing level of complexity be dissected into specific histology, grade of tumor, proliferation status, expression of tumor specific antigens, mRNA expression patterns, etc. These phenotypic endpoints may be in addition dynamic – they change with time. In an overall pool a buffering of the genetic effect is expected and it is the *shifting environment* that may, for only some specific time point of the disease, cause a given set of SNPs to acquire phenotype of relevance for its aetiology.

EXAMPLES OF GENES WITH RELEVANCE FOR STUDIES OF GENETIC SUSCEPTIBILITY

Carcinogen-DNA interactions: molecular archeology of carcinogenesis

Environmental factors, such as constituents of cigarette smoke, exposure to pesticides (pseudo- and anti-oestrogenic agents) have been implied in the primary aetiology of cancer. The enzymes involved in their metabolism are interesting targets in studies for polymorphisms leading to interindividual differences in the activation/deactivation of these substances creating a unique genetic make-up for every individual (14-16). The interactions between the genetic make-up of an individual with these environmental factors may be reflected in the type and frequency of the observed somatic mutations in various types of cancer. Mutations in the *TP53* gene for instance are the most common alterations in human cancer and TP53 is at the crossroads of a network of cellular pathways including cell cycle check-points, DNA repair, chromosomal segregation and apoptosis. A recent review of all studied mutations in the *TP53* gene in cancer patients pointed out different characteristic patterns of mutations, strongly suggesting the role of country-to-country differences in diet, life style or local environ-

ment (17). Furthermore, there are several examples of specific carcinogen exposures that are linked to cancers via *TP53* mutational mechanism such as ultraviolet light exposure, dietary aflatoxin B1 and cigarette smoking as well as combined exposures such as alcohol drinking and cigarette smoking (18). These exposures have been associated with a specific pattern of *TP53* mutations, leading to the idea that the analysis of the mutations may reveal the environmental cause back in time, thus using mutational spectra in tumour suppressor genes for such "molecular archeology" (19). A growing database of *TP53* mutations (http//www.iarc.fr/p53/homepage.html) will enable to precise such studies for mutational spectrum in attempt to search for gene-environment interactions.

Interactions between genetic variants of xenobiotic metabolising enzymes, glutathione S transferases, and somatic occurrence of mutations in the cell cycle regulator TP53

Up to 45% of the overall cancer risk is related to environmental factors and for some cancers it is even higher. Cell cycle regulation defects, on the other hand, have been well documented as molecular mechanisms leading to malignancy. Interactions between genetic variants of xenobiotic metabolising enzymes (glutathione S transferases (GSTs)) and somatic occurrence of mutations in the cell cycle regulator *TP53* were studied in our laboratory. A multiplex PCR based method for a rapid and high throughput genotype analysis of all three *GSTM1, GSTT1* and *GSTP1* genes in a single tube was developed. In breast cancer patients, carriers of the G allele of *GSTP1* had more frequently mutations in the *TP53* gene in their tumour (38%), compared to patients with the AA genotype (21%) (p=0.055) (20). These data have been confirmed in another series of breast cancer patients with locally advanced breast cancer where the GG genotype has been found a lot more frequently among the patients than among control individuals (p<0.0001) (unpublished). Furthermore, we analysed a pentanucleotide repeat in the 5' flanking area of *GSTP1* reported in GeneBank and the literature as (AAAAT) repeat. We could demonstrate that the (ATAAA)n repeat is further degenerated. The analysis of 196 healthy control individuals revealed 14 different alleles with inserts like AACAC, AAATT, AATTT, AATAT in combination with different number of repeats (21). Our report is the first observation of an extensively polymorphic area in human *GSTP1*. These findings are interesting in the light of a very recently reported pentanucleotide repeat in a PIG gene as the first TP53-responsive element found to be polymorphic (22).

CYP (Cytochrome P450)

Cytochrome P450 are mixed function oxidases which have the ability to incorporate a singlet oxygen into the

rings of hydrophobic molecules, thus priming them for further hydroxylation, acetylation or glutathione conjugation. Cytochromes P450 can thus metabolise both endogenous hydrophobic molecules (steroid hormones, cholesterol, oestrogens) and exogenous pseudo-oestrogens and other xenobiotics (7). The Human Genome project revealed the existence of 58 different genes of the human CYP superfamily, where families 1-3 are involved in the metabolism of drugs and xenobiotics (1). The majority of these genes are polymorphic and a number of functionally important variants have been described (8,9). The more extensively studied polymorphisms are summarised at http://www.imm.ki.se/cypalleles. Most of CYPs are highly inducible and may therefore be used as markers of exposure (10). For example, a vast characterisation of various CYP forms in seals from various reference sites in the Baltic sea has been reported recently (11). However, different classes of chemicals may induce the same CYP, as well as a single toxicant may induce various CYPs due to a network of "orphan" nuclear receptors, such as CAR, PXR and PPAR. These in turn cross talk to other members of intracellular signalling pathways, including those of cytokines and growth factors (12,13), thus mediating the link between the outer and inner milieu and regulation of the cell.

Xeno-oestrogens

A large amount of evidence has implicated hormones and other compounds with oestrogen activity in the pathogenesis of certain endocrine cancers, particularly breast cancer. Widely dispersed hormone-like chemicals, capable of disrupting the endocrine system and interfering with proliferation have been described. Compounds such as pesticides, some polychlorinated biphenyls and the plastic ingredient bisphenol-A have been shown to interfere with human reproduction and hormonal regulation (23,24). The levels of these foreign compounds as well as the levels of endogenous oestradiol may influence the risk of breast cancer (25). Endogenous oestradiol is synthesised in the ovarian theca cells of premenopausal women or in the stromal adipose cells of the breast of postmenopausal women and minor quantities in peripheral tissue. These cells, as well as breast cancer tissue, express all the necessary enzymes for this synthesis, the majority of them being cytochromes P450: *CYP17, CYP11a, CYP19*, hydroxysteroid hydrogenase, steroid sulphatase as well as enzymes further hydroxylating oestradiol such as *CYP1A1, CYP3A4, CYP1B* (reviewed in 26). Polymorphisms in these enzymes may have a possible role in the link between environmental oestrogens and hormone-like substances and the risk of breast cancer (reviewed in 27). Extensive work in our laboratory is devoted to characterising these polymorphisms and studying their influence on the mRNA expression and metabolic status of both control individuals and breast cancer patients (28,29).

Methods for detection of environmental endocrine disrupters

Several methodological approaches have been suggested recently to identify compounds able to disrupt normal endocrine homeostasis. Chemicals could be tested to determine their ability to displace oestrogen from its complex with hER[salpha] (human oestrogen receptor alpha) and to modulate the interaction between hER[salpha] and SRC-1 (the steroid receptor coactivator) (30). Overall oestrogen receptor related transcriptional activation in yeast (31) or in human breast cancer MCF-7 cell lines (32) have also been used to monitor the hormone disrupting potential of environmental chemicals using marker genes such as the transforming growth factor beta3 (TGFbeta3) or monoamine oxidase A. Another approach uses a battery of reporter plasmid vectors that contain firefly luciferase gene under hormone inducible control with enhancer elements responding to oestrogen, androgen, or rethinoic acid (33). Stable transfection of these reporter plasmids in ovarian carcinoma (BG-1) cell line has been used to demonstrate the potential of this bioassay to screen for known and identify unknown xeno-oestrogens (32). Several approaches have come from the field of environmental analytical chemistry, such as mass spectrometry based identification of steroid hormones in environmental matrices (33).

The development of techniques to identify natural and synthetic oestrogens in biological fluids as well as in the environment will enable to identify substances with hormonally active properties. Further genetic, biochemical metabolic analyses and exposure assessments are needed to verify the potential risk to humans.

CURRENT HIGH THROUGHPUT PLATFORMS FOR SNP ANALYSIS

The tremendous developments of identification of genetic variants in both the chemistry and bioinformatics press small and middle size academic units like us to improve and intensify by orders of magnitude the existing methods for genotyping. We have developed and published several methods for intensifying the genotyping process – of minisatellite repeats (34), multiplex PCR based analysis of polymorhisms in the glutathione-S-transferase genes (20) and a universally applied single track sequencing (SSR) (35). By these and other conventional methods we have carried out a total number of 9865 genotyping reactions (recently summarised and submitted for the European GENSUT consortium meta-analysis study, IARC, Lyon) and created a genotype database for numerous polymorphisms in genes like GST, CYP19, CYP17, epoxide hydrolase, NAT1 and CYP2D6. We have studied the existing possibilities for high throughput genotyping (recently reviewed in 36). The results of our survey on existing methods and platforms for SNP analysis are

based on three criteria: 1. The theoretical soundness of the method, 2. robustness and throughput per today and 3. availability of support and infrastructure around the platform itself.

Theoretical analysis of existing methods

The process of mutation analysis is formally divided into two steps: 1. *identification* of mutations according to the physical or enzymatic principle used to reflect the change in the DNA primary structure and 2. *visualization* of the detection products, which involves ways of making this change visible – e.g. labeling and allele separation strategies. The various approaches for allele discrimination are formally systematically divided into 1. *enzymatic* approaches, where the properties of different enzymes to discriminate between nucleotides are used (restriction enzymes type II, Cleavase and Resolvase, DNA polymerase, ligase), 2. *electrophoretic* methods, where the allele discrimination is based on difference in mobility in polymeric gels or capillaries (Single and double stranded conformation assays, heteroduplex analysis and DNA sequencing), 3. *solid phase* determination of allelic variants, including high density oligonucleotide arrays for hybridisation analysis, minisequencing primer extension analysis and fiberoptic DNA sensor arrays, 4. *chromatographic* methods – Denaturing High Performance Liquid Chromatography (DHPLC), 5. other *physical* methods of discrimination of allelic variants like mass spectrometry (mass and charge), or fluorescence exchange based techniques, and 6. *in silico* – high throughput analysis of EST data.

Robustness and throughput potential

Of these approaches, array based formats for solid phase determination seems to provide the necessary throughput and robustness. Of the array format, the Affymetrix chip, although with low sensitivity, seemed to be the only available chip when our survey was initiated (2000). The much more sound primer extension assay seemed to be hampered by the unavailability of a 5'-3' chip for a large number of interrogations. This technical problem has been circumvented today and is provided by several existing platforms (for a recent review see 37). Current producers of platforms for high throughput SNP analysis are Affymetrix (based on probe hybridisation), Sequenome (based on mass spectrometry), Invader (FLAP endonuclease), Orchid BioSciences (primer extension reaction), ABI (Taqman allele specific PCR), and Rolling circle amplification (Amersham Pharmacia). These platforms are briefly described below:

- *Affymetrix (based on probe hybridisation)* delivers high throughput, relatively low sensitivity, poor discrimination of deletions, repetitive sequences, relative dependence on the company for infrastructure.

- *Sequenome (based on mass spectrometry)* has a high potential, sound chemistry, however requires primer extension anyway, high operative costs. The system has been quite successful, involved in large projects with NIH, entered collaboration with Incyte.
- *Invader (FLAP endonuclease)*. Interesting chemistry, possibility for PCR free SNP detection, however still poorly documented and with no robustness. Requires large amount of DNA. Large project – the Japan Human Genome Project.
- *Orchid BioSciences (primer extension reaction)*. The most sound chemistry of all large throughput assays, available at three different formats, suitable for different study design – few SNP – large sample size, versus large number of SNPs on fewer individuals. Of the few who had published real results – 60000 genotypes in three different ethnic populations. Available primer sets and assays for these genotypes. Low operative cost – no labelling of the probes. Flexibility – the same probes can be used for homogenous minisequencing reactions, electrophoretically separated on capillary instruments (i.e. *ABI310* or *MegaBace*).
- *ABI (Taqman allele specific PCR)*. Large throughput and relatively high sensitivity. Availability of the instrument. High operational cost increasing with the sample size due to the usage of expensive labelled probes.
- *Rolling circle amplification (Amersham Pharmacia).* Allows signal amplification. Based on large size circularised oligonucleotides, demands specific expertise in primer design. Difficult to address to solid formats due to steric hindrance of the circular structures.

SUMMARY

Large scale genotyping studies will enable to pinpoint the SNPs which may be relevant for further studies in relation to susceptibility markers, clinical and histopathological characteristics of disease, prognosis and response to treatment. To be able to achieve these goals we need to evaluate and introduce methods for high throughput SNP analysis. The problems of identifying important variation by statistical methods will be substantial. It is therefore important that studies of critical biological expression of genetic variation work in tandem with epidemiological studies in attempts to identify relevant genetic variants that are likely to be associated with disease. We hope to be able to establish genetically predisposed variations in biologically relevant pathways involved in given environmental and life-style characteristics, exposure patterns, clinical and histopathological phenotypes as well as response to treatment. Although this approach of the "most obvious" relevant pathways enables to discover relevant genetic variants that may otherwise be lost in the majority, it has its drawbacks as it may

lead to overestimation of the effect of a given variant or a given pathway. The somewhat surprisingly large number of SNPs in the human genome and the large proportion of those expected to be functional starts inevitable discussion about the buffering effect of the pool of phenotypes in terms of both predisposition and clinical presentation. Therefore it is necessary to validate the results of our along-the-biological-pathway analysis in the picture of the overall genome SNP profile. The final epidemiological confirmation will most likely come from high throughput genotyping in large prospective studies, which with the large FUGE initiative in Norway will become reality both in terms of cost and technical availability.

ACKNOWLEDGEMENTS

REFERENCES

1. Ingelman-Sundberg M. Genetic susceptibility to adverse effects of drugs and environmental toxicants. The role of the CYP family of enzymes. *Mutat Res* 2001; **482**: 11-9.
2. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001; **29**: 229-3.
3. Terwilliger JD. On the resolution and feasibility of genome scanning approaches. *Adv Genet* 2001; **42**: 351-91.
4. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **31**: 599-603.
5. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001; **17**: 502-10.
6. Weiss KM, Clark AG. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 2002; **18**: 19-24.
7. Nedelcheva V, Gut I. P450 in the rat and man: methods of investigation, substrate specificities and relevance to cancer. *Xenobiotica* 1994; **24**: 1151-75.
8. De Vivo I, Hankinson SE, Li L, Colditz GA, Hunter DJ. Association of CYP1B1 polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2002; **11**: 489-92.
9. Zanger UM, Fischer J, Raimundo S, Stuven T, Evert BO, Schwab M, Eichelbaum M. Comprehensive analysis of the genetic factors determining expression and function of hepatic CYP2D6. *Pharmacogenetics* 2001; **11**: 573-85.
10. Soucek P, Gut I. Cytochromes P-450 in rats: structures, functions, properties and relevant human forms. *Xenobiotica* 1992; **22**: 83-103.
11. Nyman M, Raunio H, Taavitsainen P, Pelkonen O. Characterization of xenobiotic-metabolizing cytochrome P450 (CYP) forms in ringed and grey seals from the Baltic Sea and reference sites. *Comp Biochem Physiol [C]* 2001; **128**: 99-112.
12. Waxman DJ. Protein P450 gene induction by structurally diverse xenochemicals: central role of nuclear receptors CAR, PXR, and PPAR. *Arch Biochem Biophys* 1999; **369**: 11-23.
13. Smirlis D, Muangmoonchai R, Edwards M, Phillips IR, Shephard EA. Orphan receptor promiscuity in the induction of cytochromes p450 by xenobiotics. *J Biol Chem* 2001; **20**: 12822-6.
14. Rannung A, Alexandrie AK, Persson I, Ingelman-Sundberg M. Genetic polymorphism of cytochrome P4501A1, 2D6 and 2E1: regulation and toxicological significance. *J Occup Environ Med* 1995; **37**: 25-36.
15. Meyer UA. Polymorphisms of human acetyltransferases. *Environ Health Perspect* 1994; **102**: 213-6.
16. Pemble S, Schroeder KR, Spencer SR, Meyer DJ, Hallier E, Bolt H, Ketterer B, Taylor JB. Human gluthatione S-transferase Theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. *Biochem J* 1994; **300**: 271-6.
17. Sarhanis P, Redman C, Perrett C, Brannigan K, Clayton RN, Hand P, Musgrove C, Suarez V, Jones P, Fryer AA, Farrell, WE, Strange RC. Epithelial ovarian cancer: influence of polymorphism at the Gluthatione-S-transferase GSTM1 and GSTT1 loci on p53 expression. *Br J Cancer* 1996; **74**: 1757-61.
18. Shields PG, Harris CC. Cancer risk and low-penetrance susceptibility genes in gene-environment interactions. *J Clin Oncol* 2000; **18**: 2309-15.
19. Hussain SP, Hollstein MH, Harris CC. p53 tumor suppressor gene: at the crossroads of molecular carcinogenesis, molecular epidemiology, and human risk assessment. *Ann N Y Acad Sci* 2000; **919**: 79-85.

20. Nedelcheva Kristensen V, Andersen TI, Erikstein B, Nesland J, Børresen-Dale AL. Single tube method for genotyping of GSTM1, T1 and GSTP1. Relation of genotypes to P53 tumour status and clinicopathological variables in breast cancer patients. *Pharmacogenetics* 1998; **8**: 441-8.

21. Nedelcheva Kristensen V, Børresen-Dale A-L, Kristensen T. Hypervariable area in the 5' flanking region of GSTP1, previously reported as a minisatellite ATAAA repeat. *Hum Mutat* 2001; **17**: 238-9.

22. Contente A, Dittmer A, Koch MC, Roth R, Dobbelstein MA. polymorphic microsatellite that mediates induction of PIG3 by p53. *Nature Genet* 2002; **30**: 315-20.

23. Wolf MS, Toniolo PG. Environmental organochlorine exposures as a potential ethiological factor in breast cancer. *Environ Health Perspect* 1995; **103**: 141-5.

24. Bradlow HL, Devis DL, Lin G, Sepkovic D, Tiwari R. Effects of pesticides on the ratio of 16OH hydroxy-estrone: a biologic marker of breast cancer risk. *Environ Health Perspect* 1995; **103**: 147-9.

25. Pike MC, Spicer DV, Dahmoush L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. *Epidemiol Rev* 1993; **15**: 17-35.

26. Kristensen VN, Kure EH, Erikstein B, Harada N, Borresen-Dale A. Genetic susceptibility and environmental estrogen-like compounds. *Mutat Res* 2001; **482**: 77-8.

27. Kristensen VN, Borresen-Dale AL. Molecular epidemiology of breast cancer: genetic variation in steroid hormone metabolism. *Mutat Res* 2000; **462**: 323-33.

28. Kristensen VN, Andersen TI, Lindblom A, Erikstein B, Magnus P, Borresen-Dale AL. A rare CYP19 (aromatase) variant may increase the risk of breast cancer. *Pharmacogenetics* 1998; **8**: 43-8.

29. Kristensen VN, Harada N, Yoshimura N, Haraldsen E, Lonning PE, Erikstein B, Karesen R, Kristensen T, Borresen-Dale AL. Genetic variants of CYP19 (aromatase) and breast cancer risk. *Oncogene* 2000; **19**: 1329-33.

30. Lascombe I, Beffa D, Ruegg U, Tarradellas J, Wahli W. Estrogenic activity assessment of environmental chemicals using in vitro assays: identification of two new estrogenic compounds. *Environ Health Perspect* 2000; **108**: 621-9.

31. Jorgensen M, Vendelbo B, Skakkebaek NE, Leffers H. Assaying estrogenicity by quantitating the expression levels of endogenous estrogen-regulated genes. *Environ Health Perspect* 2000; **108**: 403-12.

32. Rogers JM, Denison MS. Recombinant cell bioassays for endocrine disruptors: development of a stably transfected human ovarian cell line for the detection of estrogenic and anti-estrogenic chemicals. *In Vitr Mol Toxicol* 2000; **13**: 67-82.

33. Croley TR, Hughes RJ, Koenig BG, Metcalfe CD, March RE. Mass spectrometry applied to the analysis of estrogens in the environment. *Rapid Commun Mass Spectrom* 2000; **14**: 1087-93.

34. Nedelcheva Kristensen V, Børresen-Dale AL. Improved electrophoretic separation of polymorphic short tandem repeats using bis-benzimide. *Biotechniques* 1997; **23**: 634.

35. Kristensen T, Nedelcheva Kristensen V, Børresen-Dale AL. High throughput screening for known mutations by automated analysis of single sequencing reactions (SSR). *BioTechniques* 1998; **24**: 832-5.

36. Nedelcheva Kristensen V, Kelefiotis D, Kristensen T, Børresen-Dale AL. High throughput methods for detection of genetic polymorphisms. *BioTechniques* 2001; **30**: 318-32.

37. Syvanen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2001; **2**: 930-42.