

SNPs/Pools: A methodology for the identification of relevant SNPs in breast cancer epidemiology

SUSANA N. SILVA¹, DANIELA GUERREIRO¹, MAFALDA GOMES¹, ANA PAULA AZEVEDO^{2,3},
GUILHERME BEZERRA DE CASTRO⁴, JOSÉ RUEFF¹ and JORGE F. GASPAR¹

Departments of ¹Genetics and ²Biochemistry, Faculty of Medical Sciences, Universidade Nova de Lisboa, Lisbon;

³Department of Clinical Pathology, Hospital São Francisco Xavier, Lisbon, Portugal;

⁴Molecular Cancer Research Center, Ed. Nacional Palacius, Cuiaba, Brazil

Received June 8, 2011; Accepted August 11, 2011

DOI: 10.3892/or.2011.1523

Abstract. The identification of allelic variants of human genes is of great importance when assessing genetic susceptibility. The emerging role of genetic polymorphisms in association studies has created the need for high throughput genotyping methodologies that allow a more rapid identification of relevant polymorphisms related to individual cancer risk enabling the extension to large-scale association studies. DNA pooling methodology may be of great importance considering the cost, time and labor that are involved in large-scale genotyping analysis carried out on individual samples. Alternatively, when using pooled samples which are made up of DNA from many individuals treated as a single sample, these factors are decrease drastically. In this way, the use of DNA pooling methodology, as a pre-selection tool, allows the identification of the most relevant polymorphisms to be studied, facilitating the estimation of the allelic frequency of each SNP in different populations. The present study initially aimed to validate the DNA pooling approach for the identification of genetic polymorphisms potentially associated with individual cancer risk generating pools with known allelic frequencies and using studies ongoing in the laboratory. Finally, our main aim was to test the accuracy of the pooled DNA analysis comparing the results of the allelic frequencies determined using pooled samples with the allelic frequency previously estimated by individual genotyping and previously published. In order to analyze the possibility of establishing differences between populations, we created DNA

pools using a Portuguese control population, a breast cancer population and a Xavante Indian population characterized by a total absence of breast cancer cases. The pools were firstly created with known allelic frequencies, previously determined by individual genotyping, and, latter, randomly incremented in sample size to 200 patients and controls. Our results showed that the DNA pooling approach was a useful tool for the analysis of allelic distribution in the different populations studied. In conclusion, our results showed that this methodology can be applied as an effective approach to identify SNPs of importance in genetic susceptibility to disease which may be used in association studies conducted subsequently by individual genotyping.

Introduction

Single nucleotide polymorphisms (SNPs) occur frequently throughout the human genome and are the most common type of markers used in genetic analysis (1,2). They are useful for epidemiologic studies to evaluate the role of genetic variants in the aetiology of human disease (3). The identification of allelic variants of human genes is thus fundamental to the assessment of genetic susceptibility in an individual genetic analysis. Consequently, more association studies are necessary to identify, in candidate genes, alleles that may confer small increments in individual susceptibility to disease (4).

The case control studies carried out to date have shown that SNPs increase the disease risk by a factor, normally, less than two. On the other hand, it is known that the predictability of the individual risk could be increased when the effect assigned to multiple polymorphisms is evaluated (5), allowing the development of preventive strategies at a population level. Accordingly, the goal of genetic analysis in human diseases is i) the identification of allelic variants of human genes, ii) determination of the frequency of these variants in different populations or ethnic groups (due to heterogeneous genetic background), iii) identification of diseases influenced by these variant and iv) the assessment of the magnitude of the associated risk (2,3).

The emerging role of genetic polymorphisms in association studies has created the need for high throughput genotyping methodologies that allow a more rapid identification of relevant polymorphisms related to individual cancer risk in large-scale association studies.

Correspondence to: Dr Susana N. Silva, Department of Genetics, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Rua da Junqueira 100, P-1349-008 Lisbon, Portugal
E-mail: snsilva.gene@fcm.unl.pt

Abbreviations: SNP, single nucleotide polymorphism; RFLP, restriction fragment length polymorphisms; MALDI-TOF, matrix-assisted laser desorption/ionization-time of flight; dHPLC, denaturing high-performance liquid chromatography; CI, confidence interval; OR, odds ratio

Key words: SNPs, DNA pools, breast cancer

Table I. Description of all pooled samples constructed.

ID pools	DNA samples in the pool (n)	DNA sample concentration	Population
Pool_A	20	2.5 ng/ μ l	Control
Pool_B	20		Control
Pool_C	20		Control
Pool_D	20		Control
Pool_E	50	3.5 ng/ μ l	Xavante Indian
Pool_F	100		Xavante Indian
Pool_G	178		Xavante Indian
Pool_H	200	2.5 ng/ μ l	Control
Pool_I	200		Breast cancer patients

One way to address the high cost, time and labor involved in large-scale genotyping is to carry out analyses, not on individual samples but, in pooled samples made up of DNA from many individuals (6) to be treated as a single sample (2,7-9). DNA pooling when applied to case-control studies may be used to select disease-associated variants which can be followed by individual genotyping using standard methodologies at the case-control level (9,10). In this way, and in order to validate this methodology towards its application for the identification of genetic polymorphisms potentially associated with individual cancer risk, pooled samples have been constructed with known allelic frequencies and used in studies ongoing in our laboratory (11).

Notably, the idea of using pooled DNA samples to reduce the burden of genotyping is not new. It was first suggested, in the context of case-control studies, by Arnheim *et al* (12) in an insulin-dependent diabetes mellitus study. Since then, a number of genotyping methods have been used for pooled DNA analysis, including gel-based analysis of restriction enzyme digests (RFLP), real-time polymerase chain reaction (TaqMan), matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectroscopy, denaturing high-performance liquid chromatography (dHPLC), SnaPshot and pyrosequencing, and all have provided comparable accuracy (7,10,13-16).

We previously tested the accuracy of pooled DNA analysis using RFLP analysis as a method for the estimation of allelic frequencies, comparing these with the allelic frequencies previously estimated by individual genotyping (17-19). In the present study we studied several polymorphic genes (*ERCC2* Lys751Gln, *OGG1* Ser326Cys, *XRCC1* Arg194Trp and Arg399Gln) in order to validate the reproducibility of this methodology. The main aim was to generate DNA pools using a Portuguese control population, a breast cancer patient population and a Xavante Indian population and correlate the allelic frequencies determined by individual genotyping with those obtained using pooling methodologies.

Materials and methods

DNA sample quantification and pool construction. The studied DNA samples are routinely used in our laboratory

for breast cancer association studies, including DNA samples from the Xavante Indian population which is characterized by an absolute absence of breast cancer (20). All DNA samples were extracted from 250 μ l of whole blood using a commercially available kit according to the manufacturer's instructions (QIAamp DNA extraction kit; Qiagen, Hilden, Germany).

Determination of the concentration of all DNA samples used for the construction of the pools was carried out using the PicoGreen dsDNA quantification reagent (Molecular Probes, Eugene, OR, USA) in an Anthos Zenyth 3100 (Anthos-Labtec Instruments GmbH, Austria). The linearity of the method was verified for the high-range standard curve (2 μ g/ml of Lambda DNA standard) according to the manufacturer's recommendations prior to determining the DNA concentration. Whenever individual sample concentrations showed standard errors >10%, they were re-quantified or excluded from the pool construction.

After quantification, all samples were diluted to the same concentration (Table I), and subsequently pooled using equal volumes of each sample. To verify the linearity of this method we used four pools (Pool_A to Pool_D) (Table I) constructed with 20 different DNA samples creating the different expected allelic frequencies (90, 80, 70 and 50% from the common allele, respectively) of the *ERCC2* Lys751Gln gene polymorphism. The same pools were then used to verify whether the allelic frequencies obtained using this methodology were in accordance with the ones obtained by individual genotyping of the *OGG1* Ser326Cys gene polymorphism. After the initial optimization, the pool sizes were increased to 50 and 100, with random individual DNA samples from the Xavante population (Pool_E and Pool_F) and another pool (Pool_G) was constructed using all DNA samples of the breast cancer-resistant population, and finally to 200 (Pool_H and Pool_I) corresponding to different populations currently used in our breast cancer association studies, a control and a case population, respectively (Table I). These pools were analyzed by PCR-RFLP for different polymorphic genes previously genotyped and with known allelic frequencies to ensure the accuracy of this approach.

Individual genotyping. The individual genotyping of *ERCC2* (Lys751Gln), *OGG1* (Ser326Cys) and *XRCC1* (Arg194Trp and Arg399Gln) was determined by PCR-RFLP. The primers and

Table II. Relationship between the relative peak heights and the allelic frequency in pools constructed to test method linearity.

Pool ID	n	<i>ERCC2</i> Lys751			<i>OGGI</i> Ser326		
		Individual PCR/RFLP analysis	DNA pool (\pm ST)	Δ	Individual PCR/RFLP analysis	DNA pool (\pm DP)	Δ
Pool_A	20	90.0	81.4 \pm 5.1	0.086	72.5	75.0 \pm 0.4	-0.025
Pool_B	20	80.0	78.2 \pm 1.8	0.018	70.0	77.3 \pm 1.6	-0.073
Pool_C	20	70.0	69.4 \pm 7.1	0.006	75.0	80.2 \pm 2.2	-0.052
Pool_D	20	50.0	49.7 \pm 5.4	0.003	77.5	70.8 \pm 5.3	0.067

Δ , the difference between allelic frequencies obtained by individual genotyping and frequencies obtained by the pooling methodology.

PCR conditions for the polymorphic sites of these genes have been published elsewhere (17-19). For all of them the nucleotide polymorphisms resulted in either a gain or loss of restriction site, which therefore allowed the common and variant alleles to be discriminated by RFLP after appropriate restriction enzyme digestion. All genotype determinations were carried out twice in independent experiments, and inconclusive samples were re-analyzed.

Allelic frequencies of the polymorphisms determined by pooling analysis. The approach chosen to develop the pooling methodology was based on PCR-RFLP. Each SNP can be exploited to create PCR fragments of differing size, where the PCR product is digested with a restriction enzyme endonuclease that cleaves to a fragment from one bi-allelic SNP but not the other. This method generates specific fragments to the correspondent SNP alleles, which can be detected by conventional electrophoresis on gels.

We used the ABI 377 automated DNA sequencer and GeneScan software (Applied Biosystems, Foster City, CA, USA) to quantify allele amplification of polymorphic genes (10) using 5' fluorescently labeled (with 6-FAM) reverse PCR primers. All conditions were similar to the ones described for individual genotyping. Allelic frequencies were estimated from the relative values of the peak heights corresponding to each allele detected in the pooled sample, as described by Shaw *et al* (8).

Results

The linearity of the pools was previously verified through the construction of DNA pools with known allelic frequencies as described by Gomes *et al* (11). First, we validated this methodology towards its application for identification of genetic polymorphisms potentially associated with individual cancer risk. We generated seven different DNA pools with 50 DNA samples from a population previously genotyped for the *MTHFR* (C677T) polymorphism creating different amounts of each allele (10, 20, 40, 50, 60, 80, 90% for common allele). The results obtained (data not shown) demonstrated the reproducibility of this methodology when the allelic frequencies between individual and pooled genotyping were compared. Subsequently, in order to assess the accuracy of this methodology we tested the pooling approach for several SNPs

(*ERCC2* Lys751Gln, *OGGI* Ser326Cys, *XRCC1* Arg194Trp and Arg399Gln) across pools with known allelic frequencies. This study was carried out to test whether we could use pooling methodology as a screening tool to determine the allelic frequency of several polymorphic genes in the different pools constructed, when comparing with previous individual genotyping.

First of all, we constructed pools with known allelic frequency for the *ERCC2* Lys751Gln and *OGGI* Ser326Cys gene polymorphisms (Pool_A to Pool_D). To assess the reproducibility of this technique a cut-off value of $\pm 5\%$ ($\Delta < 0.050$) was established. In other words, the differences obtained between both genotyping methods should not exceed 5%. This value allows for supporting some errors inherent to the pool preparation, as well as some errors due to the endonuclease restriction process. In this way, and following this standardization, the results obtained between individual genotyping and the pooled samples were very heterogeneous.

When allelic frequencies obtained by individual genotyping using PCR/RFLP were compared to those obtained through the pooled samples for the *ERCC2* Lys751Gln polymorphism, just one of the pooled samples showed a higher value than the cut-off value set (Pool_A with $\Delta = 0.086$) (Table II). For the SNP in the *OGGI* gene, in this first set of pools (Table II), the cut-off values varied extensively; most of them surpassing the cut-off value set for this study ($\Delta < 0.050$), which may have been related to a weak cleavage reaction corresponding to the existence of uncut PCR products that did not suffer endonuclease restriction causing significant differences in allelic discrimination.

From the results obtained, which revealed the linearity of the pooling methodology, the sample size of each pool was increased using new pools, and other SNPs in different genes were studied to test the reproducibility of the methodology. For this first increment, the Xavantes DNA samples were used. The increment in pool size revealed consistent results in allelic discrimination (Table III) when compared with previous results determined by individual genotyping, being much more evident when the allelic frequencies were compared before the increment obtained for the *OGGI* Ser326Cys polymorphism (Table II). The results obtained for the latter, in a first analysis, appeared to be much more variable, after an individual increment in the new pools, pointing to a decrease in the cut-off value ($\Delta = 0.020$ and $\Delta = 0.040$ to Pool E and F, respectively).

Table III. Relationship between the relative height peaks and the allelic frequency in pools incremented in sample size and including Xavantes Indians.

ID pools	n	<i>ERCC2</i> Lys751			<i>OGG1</i> Ser326		
		Individual PCR/RFLP analysis	DNA pool (\pm ST)	Δ	Individual PCR/RFLP analysis	DNA pool (\pm DP)	Δ
Pool_E	50	80.0	79.9 \pm 3.0	0.001	73.0	75.0 \pm 0.4	0.02
Pool_F	100	77.5	77.3 \pm 0.8	0.002	78.0	82.0 \pm 0.8	0.04

Δ , the difference between allelic frequencies obtained by individual genotyping and frequencies obtained by pooling methodology.

Table IV. Relationship between the relative height peaks and the allelic frequency in the different pools, using a case-control study as previous published.

ID pools	n	<i>XRCC1</i> Arg194			<i>XRCC1</i> Arg399		
		Individual PCR/RFLP analysis	DNA pool (\pm ST)	Δ	Individual PCR/RFLP analysis	DNA pool (\pm DP)	Δ
Pool_E ^a	50	76.0	71.3 \pm 4.4	0.047	87.0	88.3 \pm 2.4	0.013
Pool_F ^a	100	78.0	75.6 \pm 3.7	0.024	85.0	84.7 \pm 3.5	0.003
Pool_G ^a	178	74.0	74.4 \pm 2.2	0.004	83.0	78.1 \pm 5.5	0.049
Pool_H	200	91.5	86.1 \pm 1.4	0.054	63.0	63.7 \pm 1.8	0.007
Pool_I	200	93.8	89.3 \pm 0.9	0.045	67.8	67.4 \pm 2.4	0.004

Δ , the difference between allelic frequencies obtained by individual genotyping and frequencies obtained by pooling methodology. ^aXavante Indian pool.

To proceed with the validation of the pooling approach, we used a case-control study of the Portuguese population developed by us to consider the role of the *XRCC1* gene in individual susceptibility to breast cancer risk (19). We used pooling methodology to compare our previously published results with those obtained by pooled genotyping. For this, we constructed three larger pools using the population samples described in our case-control study: a case pool (Pool_I) consisting of breast cancer patients (19), a control pool (Pool_H) and a Xavante Indian pool (Pool_G). The determination of allelic frequencies of all the described pools were in accordance with the results obtained by individual genotyping.

The coefficients of variation (standard deviation/mean) for the tested pools, relative to the frequencies obtained, were 0.03 for Pool_G (Xavante Indian pool), 0.02 for Pool_H (control pool) and 0.01 for Pool_I (case pool) for the *XRCC1* Arg194Trp polymorphism, and 0.07, 0.04 and 0.03 for the same pools, respectively, for the *XRCC1* Arg399Gln polymorphism. The percentage of variation did not exceed 5% in all but one pool which emphasizes the reproducibility of this methodology. The exception was Pool_G for the Arg399Gln polymorphism which exhibited a 7% variation. These results were in accordance with those described in Table IV referring to cut-off values. However, a borderline value was obtained for Pool_H for SNP *XRCC1* Arg194Trp, which may be related to inherent errors such as partial cleavage or even laboratorial systematic errors that may occur during preparation of the pools.

Discussion

The present study demonstrated that DNA pooling methodology shows great potential for the estimation of allelic frequencies with high accuracy, a high reproducibility rate and high sensitivity. This indicates that this methodology represents a useful and efficient tool which can be used to select relevant SNPs important in association studies. Another aspect related to this technique is the low cost and reduced manpower inherent when pooled samples are used compared to analyses of individual samples; cost and manpower are increased significantly with sample size (4,7,10,21-23).

Even though there are several quantitative genotyping assays that may be applied to studies using pooled samples (primer extension, denaturing high performance liquid chromatography, cleavage at the SNP site, amplification with allele-specific primers, microarray hybridization, real-time PCR) (6,10,21,22,24), we chose a technique based on cleavage with a restriction endonuclease, as it does not require the purchasing of expensive equipment. However, using cleavage-based methods as genotyping methodology, we must ensure that cleavage reactions are complete, since any tendency towards partial cleavage results in a systematic overestimation of the allele that corresponds to the uncut PCR product (6). Nevertheless, when all methods were considered using relative quantification of allelic frequencies, previous results (6,14,25) did not attain a perfect match (100%) between the

allelic frequencies calculated from pools of individuals and frequencies determined by individual genotyping. These differences may be attributed to several factors that can contribute to this inaccuracy whichever the technology used. However, all the studies carried out to date using a pooling methodology as a screening approach, even with different genotyping technologies, have shown consistent and accurate results (2,6,7,13,14,22).

The DNA pooling approach has been increasingly used, mainly applied in the analyses of SNPs and haplotypes, in the hope that this methodology may accelerate the discovery of potential SNPs in genes related to polygenic diseases and complex phenotypes (6).

After analyzing the results obtained by individual genotyping (19) and assessing the allelic frequencies obtained for the whole set of individuals involved in the study populations, the allelic frequency obtained for the Arg194 allele of the *XRCC1* gene for the case population was 90.7%, while for the control population this frequency was 92.8%. For the same gene, the Arg399 allele revealed a frequency for the same populations of 68.0% in cases and 65.0% in controls. The results obtained showed that, overall, both populations were not significantly different in their allelic distribution, as compared with the frequencies obtained in the pooled samples, as shown in Table II.

The use of previous results obtained by a case-control study (19), allowed us to establish an overall evaluation between both genotyping methods. Data published by Silva *et al* (19), after individual analysis of each SNP included, did not show an association between SNPs and individual susceptibility to breast cancer. However, after population stratification according to menopausal status (pre-menopausal, age <44, 45 menopausal age, >45 and <54 and post-menopausal age, ≥55), the data revealed that post-menopausal women showed an increased risk of breast cancer related to specific genotypes of each SNP. Concerning the Arg194Trp polymorphism, it was observed that heterozygous individuals (Arg/Trp) with a menopausal age between 45 and 54 were at an increased risk for breast cancer [adjusted odds ratio (OR), 1.964; confidence interval (CI), 1.174-3.288; $P=0.01$] and the same effect was observed for carriers of the variant allele (Arg/Trp + Trp/Trp) (adjusted OR, 1.932; CI, 1.156-3.228; $P=0.012$). For the Arg399Gln polymorphism it was also observed that women with a post-menopausal age >55 years and homozygous for the variant form were at increased risk for this pathology (adjusted OR, 4.074; CI, 1.562-10.626; $P=0.004$) (19). However, it should be noted that the results obtained after stratification could be associated with a type I statistical error. Based on these results, it is clear that the DNA pooling approach does not allow this type of genotypic stratification as was described by Silva *et al* (19).

Considering the results shown in Table IV, it is very useful to verify the allelic distribution between different populations. Thus, Pools_E and F represent the Xavante population while Pools_H and I are healthy controls and breast cancer patients, respectively, from a Portuguese Caucasian population.

However, the accuracy of the frequency estimates may be reduced due to several factors; some intrinsic to the technique used, as the unequal representation of both alleles, that may be the result of differential PCR amplification of alleles; or the partial cleavage of PCR product that results in a systematic

overestimation of the allele that corresponds to the uncut PCR product. It is therefore possible using DNA pooling methodology to pre-select relevant SNPs to include in association studies. The ultimate purpose of these association studies is the identification of a genetic target, which accounts for the phenotype under investigation. It can also be a useful tool for identifying differences related to the genetic background of a population. The allelic frequencies obtained allowed us to compare our data with those previously published in other association studies (26-28). Comparing the differences between allelic frequencies obtained with those described can favor population characterization through the distribution of different alleles included in each study. This information can be used to identify causative genetic loci associated with complex diseases via linkage disequilibrium between two or more loci (14,29).

The DNA pooling methodology still shows some limitations when compared with individual genotyping methods. According to the genotyping platform chosen to determine allelic frequencies using DNA pools, it is possible to find variations in the accuracy of estimated frequencies, including high-throughput technology such as genotyping arrays (14,30). However, as previously mentioned, the assembly technique involves some inherent experimental errors (pipetting errors during pool construction, preferential amplification and estimation errors) that randomly influence allelic frequency determination, and, in turn, the detection of differences between cases and controls is also affected. Another limitation is related to the limit of quantitative detection that is in the range of 0.05 to 0.1 for the minor allele, which could limit the ability of pooling strategies for the discovery of rarer variants (16).

Additionally, with a pooling approach it is not possible to study sub-phenotypes, haplotypes and specific genetic models as well as undertake studies of gene-gene or gene-environment interactions. Nevertheless, one way to obtain more information concerning the differences between populations is the construction of different pools, as sub-pools from the main pooled population create homogeneous groups, which would bring together individuals in a stratified and not a random manner, according to the type of study that is to be carried out. However, the results obtained with this type of stratification may reflect only the allelic distribution for each pool, which is of importance to determine whether any significant differences exist between the populations under study. However, pooling methodology does not exclude the use of individual genotyping to develop association studies (30,31).

Nevertheless, irrespective of all limitations, DNA pooling technique provides an attractive alternative when dealing with the costs inherent to association studies. This approach is also a very useful tool for genome-wide association studies (GWAS), as it allows allelic frequency estimation, rather than specific genotypes, allowing for analysis and the classification of the likelihood of association with a disease. The development of this technique in GWAS studies allows cost savings which are essential in studies of this size (31).

In conclusion, this methodology can be applied for identifying the main genetic associations with disease and is an effective approach for the identification of SNPs of increased interest to be used in association studies conducted subse-

quently by individual genotyping. It is of great importance in the investigation of the genetic background of complex diseases.

Acknowledgements

We wish to thank Luísa Manso Oliveira, Lylliane Luz and Maria Catarina Soveral for their technical support. The Center for Research in Human Molecular Genetics (CIGMH), Projects PTDC/SAU/OSM/105572/2008 (FCT) and Fundação Calouste Gulbenkian (Grant 69405) supported our current research. The PhD grant SFRH/BD/17828/2004 from FCT is also acknowledged.

References

- Mattarucchi E, Marsoni M, Binelli G, Passi A, Lo CF, Pasquali F and Porta G: Different real time PCR approaches for the fine quantification of SNP's alleles in DNA pools: assays development, characterization and pre-validation. *J Biochem Mol Biol* 38: 555-562, 2005.
- Yang HC, Pan CC, Lin CY and Fann CS: PDA: pooled DNA analyzer. *BMC Bioinformatics* 7: 233, 2006.
- Little J, Bradley L, Bray MS, *et al*: Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 156: 300-310, 2002.
- Norton N, Williams NM, Williams HJ, *et al*: Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. *Hum Genet* 110: 471-478, 2002.
- Tabor HK, Risch NJ and Myers RM: Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 3: 391-397, 2002.
- Sham P, Bader JS, Craig I, O'Donovan MC and Owen MJ: DNA pooling: a tool for large-scale association studies. *Nat Rev Genet* 3: 862-871, 2002.
- Chowdari KV, Northup A, Pless L, *et al*: DNA pooling: a comprehensive, multi-stage association analysis of ACSL6 and SIRT5 polymorphisms in schizophrenia. *Genes Brain Behav* 6: 229-239, 2007.
- Shaw SH, Carrasquillo MM, Kashuk C, Puffenberger EG and Chakravarti A: Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. *Genome Res* 8: 111-123, 1998.
- Zuo Y, Zou G and Zhao H: Two-stage designs in case-control association analysis. *Genetics* 173: 1747-1760, 2006.
- Rollinson S, Allan JM, Law GR, *et al*: High-throughput association testing on DNA pools to identify genetic variants that confer susceptibility to acute myeloid leukemia. *Cancer Epidemiol Biomarkers Prev* 13: 795-800, 2004.
- Gomes M, Guerreiro D, Silva SN, *et al*: DNA pooled samples strategies for identification of genetic polymorphisms associated with breast cancer risk. Abstract of the 95th Annual Meeting of the American Association for Cancer Research. Orlando, FL, USA, 2004.
- Arnheim N, Strange C and 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* 82: 6970-6974, 1985.
- Le Hellard S, Ballereau SJ, Visscher PM, *et al*: SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis. *Nucleic Acids Res* 30: e74, 2002.
- Jurinke C, Oeth P and van den Boom D: MALDI-TOF mass spectrometry: a versatile tool for high-performance DNA analysis. *Mol Biotechnol* 26: 147-164, 2004.
- Tost J and Gut IG: Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications. *Clin Biochem* 38: 335-350, 2005.
- Werner M, Sych M, Herbon N, Illig T, König IR and Wjst M: Large-scale determination of SNP allele frequencies in DNA pools using MALDI-TOF mass spectrometry. *Hum Mutat* 20: 57-64, 2002.
- Silva SN, Gil OM, Oliveira VC, *et al*: Association of polymorphisms in ERCC2 gene with non-familial thyroid cancer risk. *Cancer Epidemiol Biomarkers Prev* 14: 2407-2412, 2005.
- Silva SN, Bezerra de CG, Faber A, *et al*: The role of ERCC2 polymorphisms in breast cancer risk. *Cancer Genet Cytogenet* 170: 86-88, 2006.
- Silva SN, Moita R, Azevedo AP, *et al*: Menopausal age and XRCC1 gene polymorphisms: role in breast cancer risk. *Cancer Detect Prev* 31: 303-309, 2007.
- Silva SN, Cabral MN, Bezerra de Castro G, *et al*: Breast cancer risk and polymorphisms in genes involved in metabolism of estrogens (CYP17, HSD17 β 1, COMT and MnSOD): Possible protective role of MnSOD gene polymorphism Val/Ala and Ala/Ala in women that never breast fed. *Oncol Rep* 16: 781-788, 2006.
- Gaustadnes M, Orntoft TF, Jensen JL and Tørring N: Validation of the use of DNA pools and primer extension in association studies of sporadic colorectal cancer for selection of candidate SNPs. *Hum Mutat* 27: 187-194, 2006.
- Hoogendoorn B, Norton N, Kirov G, *et al*: Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools. *Hum Genet* 107: 488-493, 2000.
- Norton N, Williams NM, O'Donovan MC and Owen MJ: DNA pooling as a tool for large-scale association studies in complex traits. *Ann Med* 36: 146-152, 2004.
- Casado-Diaz A, Cuenca-Acevedo R, Quesada JM and Dorado G: Individual single tube genotyping and DNA pooling by allele-specific PCR to uncover associations of polymorphisms with complex diseases. *Clin Chim Acta* 376: 155-162, 2007.
- Barratt BJ, Payne F, Rance HE, Nutland S, Todd JA and Clayton DG: Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design. *Ann Hum Genet* 66: 393-405, 2002.
- Huang Y, Li L and Yu L: XRCC1 Arg399Gln, Arg194Trp and Arg280His polymorphisms in breast cancer risk: a meta-analysis. *Mutagenesis* 24: 331-339, 2009.
- Li H, Ha TC and Tai BC: XRCC1 gene polymorphisms and breast cancer risk in different populations: a meta-analysis. *Breast* 18: 183-191, 2009.
- Saadat M and Ansari-Lari M: Polymorphism of XRCC1 (at codon 399) and susceptibility to breast cancer, a meta-analysis of the literature. *Breast Cancer Res Treat* 115: 137-144, 2009.
- Cardon LR and Bell JI: Association study designs for complex diseases. *Nat Rev Genet* 2: 91-99, 2001.
- Jawaid A and Sham P: Impact and quantification of the sources of error in DNA pooling designs. *Ann Hum Genet* 73: 118-124, 2009.
- Bosse Y, Bacot F, Montpetit A, *et al*: Identification of susceptibility genes for complex diseases using pooling-based genome-wide association scans. *Hum Genet* 125: 305-318, 2009.