A Comparison of DNA Pools Constructed Following Whole Genome Amplification for Two-Stage SNP Genotyping Designs

Zhen Zhen Zhao, Dale R. Nyholt, Michael R. James, Renee Mayne, Susan A. Treloar, and Grant W. Montgomery

Molecular Epidemiology and Genetic Epidemiology Laboratories, Queensland Institute of Medical Research, Brisbane, Australia

Genotyping in DNA pools reduces the cost and the time required to complete large genotyping projects. The aim of the present study was to evaluate pooling as part of a strategy for fine mapping in regions of significant linkage. Thirty-nine single nucleotide polymorphisms (SNPs) were analyzed in two genomic DNA pools of 384 individuals each and results compared with data after typing all individuals used in the pools. There were no significant differences using data from either 2 or 8 heterozygous individuals to correct frequency estimates for unequal allelic amplification. After correction, the mean difference between estimates from the genomic pool and individual allele frequencies was .033. A major limitation of the use of DNA pools is the time and effort required to carefully adjust the concentration of each individual DNA sample before mixing aliquots. Pools were also constructed by combining DNA after Multiple Displacement Amplification (MDA). The MDA pools gave similar results to pools constructed after careful DNA quantitation (mean difference from individual genotyping .040) and MDA provides a rapid method to generate pools suitable for some applications. Pools provide a rapid and cost-effective screen to eliminate SNPs that are not polymorphic in a test population and can detect minor allele frequencies as low as 1% in the pooled samples. With current levels of accuracy, pooling is best suited to an initial screen in the SNP validation process that can provide high-throughput comparisons between cases and controls to prioritize SNPs for subsequent individual genotyping.

Finding genes contributing to complex human disease remains a major challenge. Whole genome association on clinically relevant sample sizes requires massive genotyping projects (Cardon & Bell, 2001; Risch, 2000; Tabor et al., 2002). There are limitations to individual genotyping in large-scale association studies because of high cost, the large amount of DNA required and the time needed to complete such a project (Smith & Lusis, 2002). Linkage studies can define genome regions linked to disease, but lack power to detect the modest effects expected for many common complex disorders (Lander, 1996; Risch & Merikangas, 1996; Schork et al., 1998). Even when linkage is detected, the confidence interval is generally several megabases and cannot be reduced further. Association studies to locate the gene or genes contributing to the linkage also require genotyping projects of 1 to 2 million genotypes, depending on the size of the confidence interval and size of the mapping population.

DNA pooling is one approach used successfully to reduce the cost of genotyping (Sham et al., 2002). Allele frequencies are estimated in a pool of DNA samples from cases and compared with estimates from a pool of controls. This greatly reduces the number of polymerase chain reactions (PCR) and genotyping reactions that must be performed. The use of pooling introduces additional errors that must be taken into account in the design and analysis of pooling experiments (Le Hellard et al., 2002; Sham et al., 2002). As a result, one effective way to use DNA pooling might be a two-stage design where pools are used as an initial screen and markers showing association can be followed up by individual genotyping (Sham et al., 2002).

Allele frequencies of pooled DNA samples can be accurately estimated using a range of single nucleotide polymorphism (SNP) genotyping platforms (Bansal et al., 2002; Le Hellard et al., 2002; Norton et al., 2002). The Sequenom MassARRAY genomics platform is used extensively for SNP genotyping (Braun et al., 1997; Buetow et al., 2001; Little, Braun, Darnhofer-Demar, et al., 1997; Little, Braun, O'Donnell, et al., 1997), DNA pooling (Le Hellard et al., 2002; Ross et al., 2000; Werner et al., 2002), mutation detection (Higgins et al., 1997), clinical diagnostics (Jurinke et al., 1998), and quantitative gene expression (Ding & Cantor, 2003). Frequency determination in pools with the MassARRAY platform using primer extension has been found to be as

Received 23 March, 2005; accepted 24 May, 2005.

Address for correspondence: Dr Grant Montgomery, Queensland Institute of Medical Research, 300 Herston Rd, Herston, QLD 4006, Australia. E-mail: Grant.Montgomery@qimr.edu.au

https://doi.org/10.1375/twin.8.4.353 Published online by Cambridge University Press

sensitive, reproducible and cost-effective as other available technologies (Le Hellard et al., 2002).

One critical issue for pooling studies is DNA sample preparation for constructing pools. Pools must be carefully constructed after multiple steps of measuring and readjusting to normalize DNA concentrations before samples are pooled (Sham et al., 2002). To avoid introducing extra sources of variation, sample preparation must be completed by the same skilled operator or requires robotic liquid handlers for automation. With the increasing resource of SNPs in the public database, the limiting factors in a pooling study have become the availability of DNA, time required for sample handing and the allelotyping capacity. Recently methods have been developed for whole genome amplification (WGA). One important property of methods for genome amplification is the reactions produce similar yields of DNA from a wide range of starting DNA concentrations (Dean et al., 2002; Hosono et al., 2003). Combining DNA samples after WGA might produce suitable DNA pools for two-stage pooling designs with a considerable saving in labor costs.

The aims of the present study were to evaluate pooling as part of a strategy for fine mapping in regions of significant linkage by determining the accuracy of frequency estimates in pooled samples based on a large number of SNPs, to investigate WGA for DNA pool construction, and to investigate multiplex genotyping.

Materials and Methods

Study participants were recruited from the Australian population as part of a genetic study of endometriosis (Treloar et al., 2002). Study protocols were reviewed and approved by the Queensland Institute of Medical Research Human Research Ethics Committee. Participation was voluntary and each participant gave written informed consent. For the current study, DNA samples were selected from 768 individuals and used to create two pooled samples of DNA of 384 individuals each.

Generation of DNA Pools

Genomic DNA was extracted (Miller et al., 1988) from peripheral venous blood samples. DNA concentrations were measured using PicoGreen (Molecular Probes) for the quantitation of double-stranded DNA in solution on a Fluoroskan Ascent CF plate reader (Labsystems, Chicago). Concentrations of DNA samples were initially adjusted to 30 ng/μl and carefully adjusted by serial dilutions to a final concentration of 25 ng/μl ($M \pm SD = 25.19 \pm 0.55$). Individual DNA samples were tested in at least two PCR reactions to ensure samples contained high quality DNA. Two pools of 384 individuals were constructed by mixing equal amounts of adjusted DNA samples.

DNA Amplification

Genomic DNA samples were amplified by multiple displacement assay (MDA) using the GenomiPhi DNA Amplification Kit (Amersham Biosciences Limited) according to the manufacturer's protocol. Reactions were performed in 5μl volumes, which contained 10 ng genomic DNA at 25ng/μl, 2μl of sample buffer, 2.25μl of reaction buffer and 0.25μl of Phi29 DNA polymerase. The reaction conditions were 30°C for 16 hours for the strand displacement reaction using bacteriophage Phi29 DNA polymerase, and 65°C for 10 minutes to heat-inactivate the enzyme. A 1:300 dilution was made prior to mixing equal amounts from each MDA-DNA sample to form the pools. To examine the accuracy of the MDA pooling approach, two MDA pools were constructed with samples amplified from the same individuals used to construct the two genomic DNA pools.

SNP Selection

SNPs used in this study were retrieved from the Sequenom SNP database (http://www.RealSNP.com). Thirty-nine SNPs were selected from a set of assays that had been successfully typed on 96 individuals as part of the endometriosis project. Primers were designed using the Sequenom SpectroDESIGNER software (version 2.1) and synthesized by BIONEER (Korea).

SNP Genotyping

Individual SNP typing of DNA samples used to construct the pools and allele frequency estimation of DNA pool samples were completed for 39 SNP assays. All PCR and MassEXTEND reactions were conducted using standard conditions as described elsewhere (Bansal et al., 2002). Briefly, DNA (5ng for individual DNA samples and 25ng for pools), 0.1 unit of *Taq* polymerase (HotStarTaq, Qiagen, Valencia, CA), 200 μmol of each dNTP, and 50 nmol to 300 nmol of each PCR primer were mixed in a total volume of 5μl. PCR reactions were cycled with the following temperature profile: initial denaturation at 94°C for 15 minutes followed by 45 cycles of 94°C for 20 seconds, 56°C for 30 seconds, 72°C for 1 minute, and a final extension of 3 minutes at 72°C. To genotype the pools, PCR reactions were replicated four times for each assay using the same PCR conditions.

Following PCR, shrimp alkaline phosphatase (Sequenom, San Diego, CA) was added to remove excess dNTPs from the PCR products. The reactions were incubated at 37°C for 20 minutes and then for 5 minutes at 85°C for enzyme inactivation. Triple terminator mixes were used for primer extension reactions. The multiplex assays in one PCR were grouped according to the SNP-specific requirements for the termination mix and the minimum peak separation required for the MassARRAY. The final reaction volume of 9μl comprised 0.6 units of thermosequenase (Sequenom, San Diego, CA), 300 nM to 600 nM extension primer and 0.22 × termination mix. All assays were run with the same temperature

https://doi.org/10.1375/twin.8.4.353 Published online by Cambridge University Press

profile, comprising an initial denaturation at 94°C for 2 minutes followed by 55 cycles of 94°C for 5 seconds, 52 °C for 5 seconds and 72°C for 5 seconds.

Allele Frequency Determination

Allele frequencies were estimated in pooled DNA samples using our MassARRAY™ system (Sequenom, San Diego). Aliquots of the pooled samples were spotted four times for each PCR reaction onto a 384-element silicon allelotyping chip using a Nanodispenser (Samsung, Korea). To evaluate the reproducibility of the frequency estimates, four replicate PCR reactions were carried out for each assay in each pool yielding a total of 16 observations for each pool for each SNP.

Allele frequency estimates for pools need to be adjusted for any differences in the peak height for each allele observed in heterozygous individuals. To evaluate the number of heterozygous individuals to include in the correction factor, results were compared from adjustments estimated from two or eight heterozygous individuals genotyped for each assay. We obtained peak areas from SpectroTYPER™ software (Sequenom, San Diego CA) by integration of the area under the spectral peak at the expected mass of the extension product. The heterozygous individuals were selected from initial SNP validation experiments. To calculate allele frequencies using peak areas generated from mass spectra, both alleles were analyzed twice for each observation. All of the SNPs were individually genotyped for final validation.

Statistical Analysis

Allele frequencies on pooled samples were calculated from peak areas based on independent mass spectrometry measurements of up to four observations for four replicate PCR reactions (16 observations). For each of these data points, SpectroTYPER™ software was initially used based on the pool peak areas A and B of the lower- and higher-mass alleles, respectively, to obtain the pool-based allele frequency estimate $P = A/(A + B)$. In order to adjust unequal allelic amplification, two and eight heterozygotes for each assay were genotyped individually in separate genotype + area experiments. The corrected allele frequency data was obtained based on the skew correction files generated from the genotype + area experiments.

The allele frequency difference between pooled samples and individual samples were calculated from the mean of the absolute difference between corrected allele frequency estimates in pools and the expected frequency estimated from genotyping individual samples. Comparisons of allele frequencies and genotype frequencies between groups were also carried out using a chi-square test of independence with one degree of freedom.

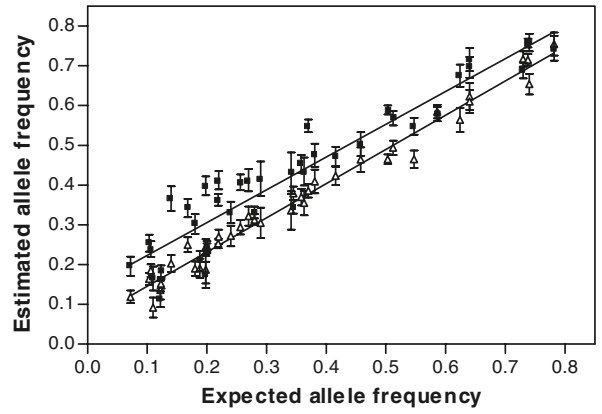

Figure 1

Estimated allele frequencies from one genomic DNA pool for uncorrected data (■ $r = .94$) and data corrected for allele bias based on data from eight heterozygous individuals (Δ $r = .99$) compared with expected allele frequencies from individual genotyping.

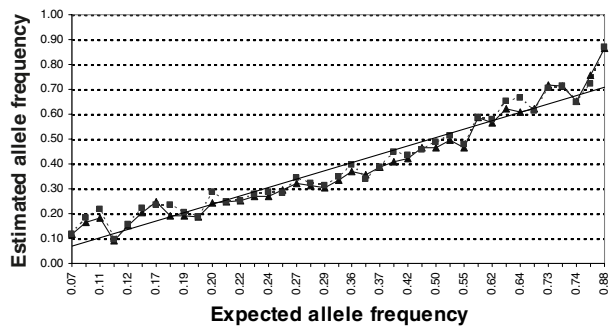

Figure 2

Estimated allele frequencies in one genomic DNA pool before and after correction for allele bias based on data from two heterozygous (■ $r = .98$) or eight heterozygous individuals (Δ $r = .99$) compared with expected allele frequencies from individual genotyping.

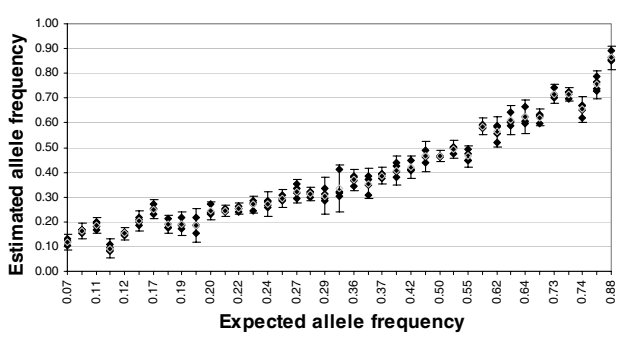

Figure 3

The mean and 95% confidence intervals for estimated allele frequencies in one genomic DNA pool for 39 SNP assays compared with expected allele frequencies from individual genotyping. Variation was estimated from four replicate PCR reactions and four replicate allele frequency estimates from each reaction.

https://doi.org/10.1375/twin.8.4.353 Published online by Cambridge University Press

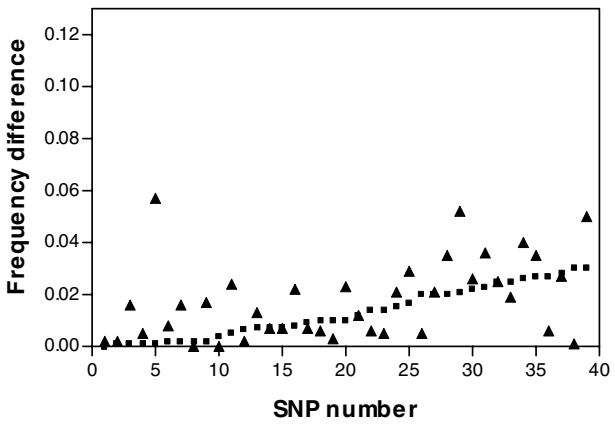

Figure 4

Absolute allele frequency differences between case and control pools estimated from 39 SNP assays (▲) compared with the differences observed from genotyping individual cases and controls used to construct the pools (■).

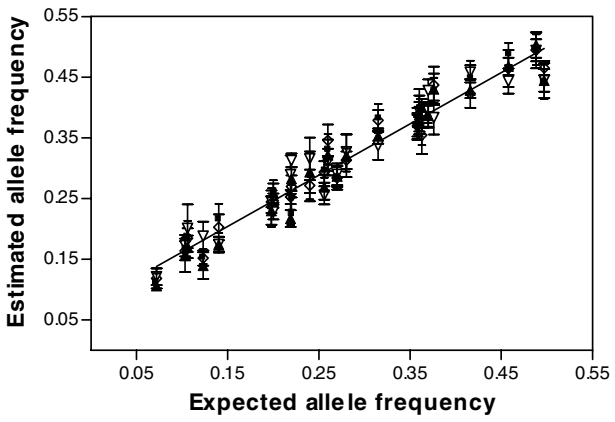

Figure 5

Allele frequencies estimated for 28 SNPs by combining PCR products for two SNPs followed by duplex SNP analysis in DNA pools (■ $r = .97$), conducting duplex PCR and analysis in DNA pools (▲ $r = .98$), and duplex PCR and analysis of DNA pools genereated following whole genome amplification (Δ $r = .97$). Duplex PCR reactions produced similar results to single-plex genotyping (◇ $r = .98$).

Results

Allele frequencies were estimated in genomic DNA pools constructed from two groups of 384 individuals and compared with observed frequencies from genotyping all 39 SNPs in the 768 individual DNA samples used to construct the pools. Genotype data were obtained for an average of 99.2% of individuals. SNPs with allele frequencies less than .05 were excluded from the analysis. Minor allele frequencies for the SNPs ranged from .065 to .497. PCRs for pooled samples were performed in quadruplicate and extension products were spotted four times for each PCR. To estimate error rates due to genotyping technical causes, one SNP was typed twice on 1920 DNAs independently at different times. The dropout rate was 0.85%. Of the 3769 complete genotypes there were five discordant genotypes which is an error frequency of 0.13%. There were no significant differences in allele frequencies for genotype data for the individual samples used to make up each DNA pool for any SNPs tested.

We tested whether the number of heterozygous individuals used to adjust allele frequency estimates would affect the accuracy of frequency estimation in pooled DNAs. We genotyped two or eight heterozygous individuals for each of 39 assays. The correction (skew) factors for each assay were calculated by the allelotyping software (SpectroTYPER) based on baseline correction and peak area calculations as peak area divided by the sum of all peak areas. The data were analyzed with each of the correction factors, and the accuracy compared with individual genotypes. In the control pool, the mean standard deviation for estimates of allele frequency was .022 for the data corrected by both correction factors (obtained from two and eight heterozygous individuals) and .024 for uncorrected data. As expected, the uncorrected data had larger errors and less accurate frequency estimation than the corrected data (Figure 1). The average correlation between estimates of allele frequency from pools and individual genotypes was .972 using corrections from two heterozygous individuals (Cor2) and .981 for using corrections from eight heterozygous individuals (Cor8). The difference between the frequency estimates from these two sets of correction factors was not significant (Figure 2).

Allele frequencies in four replicate PCRs for each pool were corrected with eight heterozygous individuals for subsequent analyses (Figure 3). The same primer extension products were spotted four times for each replicate. The mean standard deviation between the four replicates for all 39 SNPs was .017 (range .001 to .049). The accuracy of the allele frequency estimates was evaluated by comparing the frequencies observed from the pooled samples with those observed from typing the individual samples from which the pool was constituted. The mean absolute difference between the corrected allele frequency estimate in the pools and the expected frequency from the corresponding individual genotypes was .033 (range .0008 to .086). The accuracy of the allele frequency estimates was also evaluated by comparing the estimated allele frequency differences between case and control pools compared with the observed allele frequency differences from genotyping individuals comprising the pools (Figure 4). The mean absolute error in estimating the allele frequency difference between pools was .013 ± .013, ranging from .00006 to .058. The mean standard deviation for differences between the estimate for pooled samples compared with estimate from individual genotyping for the 39 SNPs was .038.

https://doi.org/10.1375/twin.8.4.353 Published online by Cambridge University Press

The cost of SNP genotyping is highly dependent on the ability to multiplex assays in the same reaction. We therefore tested duplex reactions in pooled DNA samples performed by mixing primer extension products from two uni-plex PCRs before spotting four times. The SpectroDESIGNER 2.1 software was used for multiplex design from the original 39 assays. A subset of 28 SNP assays was chosen for subsequent studies based on separation of molecular mass and the terminator nucleotides required within a multiplex group. These 28 SNPs were tested on the duplex reactions used to evaluate the pooling method. Additional pools were also constructed following WGA of DNA samples by combining equal aliquots from the amplification of individual DNA samples to replicate the original carefully constructed pools. Aliquots of the MDA reactions were pooled without prior estimation of individual DNA concentrations. Two replicates were performed for each DNA pool sample and for each set of two assays. Each extension reaction was dispensed and scanned four times. Data were corrected and analyzed using the skew correction factors for eight heterozygous individuals (Cor8).

The minor allele frequencies for these 28 SNPs ranged from .065 to .497. We compared the results of allele frequency estimation on the MassARRAY when individual PCR and extension reactions were pooled before spotting on chips, duplex PCR reactions and for pools generated from WGA. In the genomic control pool, the mean standard deviation of estimated allele frequencies observed was .022 for pooled duplex, and .020 and .021 for the duplex results on the genomic DNA pool and MDA-DNA pool respectively. Results for allele frequency estimates for individual genotyping, genomic DNA pools and MDA pools showed very good agreement (Table 1). The mean differences in allele frequency between individual genotyping and estimates from genomic (Δind-gen) and MDA (Δind-mda) pools for the 28 SNP assays were .031 (range .001 to .079) and .040 (range .001 to .122) respectively. The small increase in variability of the MDA-DNA pools might be due to variation in DNA yields or uneven amplification of alleles in individual DNA samples. Correlations between genotyping on individuals and genotyping on genomic and MDA pools were .98 and .96 respectively. Results were similar when duplex assays were used (Figure 5).

To evaluate the sensitivity of the pooling method we analyzed four SNPs that had been genotyped previously and were monomorphic in 64 individuals. In the genomic- and MDA-DNA control pool, the average minor allele frequencies for these four SNPs was .01 $\pm$ .0095 and .01 $\pm$ .0098 ($M \pm SD$) respectively. The small differences between the frequencies observed from the pooled samples and those observed from individual samples may be due to both errors in estimating frequency from the pools and/or sampling variation, since individual genotyping for these SNPs was only completed for 64 individuals from a different set of samples to those used in pool construction. By detecting allele frequencies as low as 1% in the pool, the results suggested that high throughput SNP genotyping on pools is a sensitive method to identify and exclude SNPs that would not be informative in the sample.

Discussion

A pooling strategy reduces genotyping costs and the time required to complete large genotyping projects (Bansal et al., 2002; Hoogendoorn et al., 2000; Le Hellard et al., 2002; Mohlke et al., 2002; Sham et al., 2002; Visscher & Le Hellard, 2003). However pooling results in a loss of information because of increased errors in the experimental procedure and the loss of haplotype information. Genotypes for the pools must be replicated, but there are still thirty- to fiftyfold savings in reagent costs, depending on study design (Mohlke et al., 2002; Sham et al., 2002). The MassARRAY genomics platform provides an accurate method to estimate allele frequencies in pooled DNA samples and our results are in good agreement with previous studies using this platform (Bansal et al., 2002; Le Hellard et al., 2002; Mohlke et al., 2002). We compared allele frequency estimates in two pools with results derived from genotyping the individual DNA samples used for pool construction. Assays for 39 SNPs were designed and tested on individual DNA samples and rejected only if the minor allele frequency was less than 5% or there was substantial variation between peak heights for each allele (peak height ratio greater than 75:25). Four separate PCR reactions were carried out for each pool to reduce experimental error when comparing allele frequency between pools.

The mean absolute difference between frequency estimates from individual genotypes and pooled samples was .034 similar to values of .033 to .04 reported previously using the Sequenom MassARRAY (Bansal et al., 2002; Le Hellard et al., 2002; Mohlke et al., 2002; Visscher & Le Hellard, 2003). We found allele frequency estimates must be corrected for unequal amplification of alleles, but there was no significant difference when using data from two or eight heterozygous individuals. Consequently, a modest number of individuals can be typed in a fixed panel to identify heterozygous individuals for the correction of allele frequency estimates for each SNP across the range of minor allele frequencies.

Analysing pooled samples to compare frequency estimates between groups increases Type I errors due to experimental factors including pool construction, assay quality, correction factor calculations, and individual genotype errors (Bray et al., 2001; Sham et al., 2002). The mean error in estimating allele frequency differences between the two pools of .013 was slightly higher than .009 reported for a project analyzing 16 SNPs (Mohlke et al., 2002). In our data set this was largely due to two particular SNPs. The large differ-

https://doi.org/10.1375/twin.8.4.353 Published online by Cambridge University Press

Table 1

Allele Frequencies Estimated From Individual Genotyping and DNA Pools Generated From Carefully Measured Genomic DNA Samples and Samples From Multiple Displacement Assay (MDA) Whole Genome Amplification in Two Groups of 384 Individuals (Pool 1 and Pool 2)

	Pool 1					Pool 2				
ASSAY	Individual	Genomic	Δ_{p1}ind-gen	MDA	Δ_{p1}ind-mda	Individual	Genomic	Δ_{p2}ind-gen	MDA	Δ_{p2}ind-mda
1	.297	.352	.055	.395	.098	.282	.320	.038	.327	.045
2	.355	.391	.036	.328	.027	.358	.375	.017	.367	.009
3	.063	.094	.031	.061	.002	.064	.082	.018	.080	.016
4	.343	.381	.038	.383	.040	.370	.386	.016	.427	.057
5	.306	.317	.011	.377	.071	.315	.352	.037	.336	.021
6	.415	.435	.020	.442	.027	.416	.428	.012	.458	.042
7	.221	.224	.003	.299	.078	.219	.217	.002	.289	.070
8	.218	.254	.036	.340	.122	.220	.281	.061	.312	.092
9	.196	.226	.030	.229	.033	.198	.229	.031	.244	.046
10	.210	.289	.079	.262	.052	.212	.259	.047	.227	.015
11	.104	.181	.077	.150	.046	.103	.156	.053	.172	.069
12	.146	.181	.035	.238	.092	.123	.139	.016	.188	.065
13	.115	.183	.068	.172	.057	.140	.174	.034	.174	.034
14	.140	.142	.002	.147	.007	.110	.096	.014	.133	.023
15	.116	.175	.059	.206	.090	.106	.171	.065	.201	.095
16	.065	.101	.036	.079	.014	.072	.105	.033	.121	.049
17	.286	.317	.031	.370	.084	.265	.321	.056	.310	.045
18	.290	.317	.027	.350	.060	.266	.286	.020	.293	.027
19	.458	.484	.026	.459	.001	.458	.467	.009	.443	.015
20	.410	.413	.003	.428	.018	.412	.413	.001	.428	.016
21	.371	.386	.015	.334	.037	.363	.399	.036	.385	.022
22	.117	.133	.016	.143	.026	.122	.132	.010	.125	.003
23	.381	.434	.053	.410	.029	.373	.429	.056	.384	.011
24	.491	.459	.032	.492	.001	.497	.444	.053	.444	.053
25	.478	.473	.005	.485	.007	.488	.496	.008	.494	.006
26	.385	.405	.020	.377	.008	.358	.388	.030	.393	.035
27	.270	.301	.031	.276	.006	.256	.268	.012	.255	.001
28	.218	.245	.027	.273	.055	.240	.293	.053	.315	.075

Note: Differences in allele frequency between estimates from individual genotyping and the genomic DNA pool (Δind-gen) and MDA pool (Δind-mda) are also given for pool 1 (p1) and pool 2 (p2).

https://doi.org/10.1375/twin.8.4.353 Published online by Cambridge University Press

ence for these SNPs is unlikely to be caused by pool construction or general experimental procedures since other SNPs showed expected levels of variability. Inspection of individual genotypes for these SNPs did not suggest obvious assay problems that would justify exclusion of the SNPs from the series. The outliers may be influenced by skew factor correction, but the high values demonstrate the increase in the Type I error observed for some assays using pooled samples and the necessity to verify possible differences by genotyping individual samples (Bray et al., 2001; Sham et al., 2002).

Linkage analysis in complex trait studies usually result in linkage peaks with large confidence intervals that cannot be reduced further. DNA pooling can reduce genotyping costs, but the genotyping accuracy is critical in searching for the disease gene under a linkage peak where the disease model for the trait is usually unknown. Despite limitations, DNA pooling still has several advantages for a rapid screen in these studies. SNP allele frequency information is improving rapidly (Hinds et al., 2005), but may not be available for the population under study. We recently designed and tested 326 assays from 22 genes. Despite including all SNPs with frequency information, 119 SNPs (36.5%) were not polymorphic or had very low frequencies in our population. This number of nonpolymorphic SNPs was similar to the 40% of SNPs present in dbSNP, but not observed in an independent gene-centric SNP study (Jiang et al., 2003). Initial typing in DNA pools rapidly identifies SNPs with very low frequencies in the test population and these can be excluded from further analysis. Several authors have suggested the use of two-stage experimental designs for pooling experiments to overcome the limitations and still retain the benefits of reduced costs (Mohlke et al., 2002; Sham et al., 2002). Including pools of cases and controls provides an initial screen for association that might direct attention to particular SNPs or regions to include first in individual genotyping of the test populations.

A major limitation of the use of DNA pools for these tests is the time and effort required to carefully adjust the concentration of each individual DNA sample before mixing aliquots to generate the pools. WGA using MDA uses the Phi29 DNA polymerase and random exonuclease-resistant primers in an isothermal amplification reaction (Dean et al., 2002). MDA has been used successfully to amplify DNA from clinical samples for genomic testing and provides good quality template for several different genotyping platforms (Dean et al., 2002; Hosono et al., 2003). An important property of the MDA reaction is the self-limiting nature of the reaction which generates a similar yield from widely different concentrations of starting material (Dean et al., 2002; Hosono et al., 2003). We tested pools made from MDA reaction product without further quantitation. Good agreement was found between allele frequency estimates from the MDA pools and estimates from both individual genotyping and our pools made from carefully adjusted DNA samples. These results suggest MDA provides a method to rapidly generate pools suitable for screening. It would be useful to design future studies to determine the range of DNA concentrations that could be tolerated to make MDA-DNA pools.

Typing several SNPs in the same multiplex assay can be performed easily for individual genotyping because the background noise has a small effect on the allele peak detection. Conversely, allelotyping requires a high degree of accuracy for allele frequency estimation and does not allow easy multiplexing. Special care with allelotyping is required for peak area integration to ensure reproducibility and robustness against noise and baseline errors. Limited two-plex assays were tested by either combining separate PCR products or conducting duplex assays for combinations of 28 SNPs. There was excellent agreement between results for the single and duplex assays. Preliminary experiments with four-plex assays have not been successful.

In conclusion, with current levels of accuracy, pooling is best suited to an initial screen in the SNP validation process for association studies in regions of linkage. Pools provide a rapid and cost-effective screen to detect minor allele frequencies as low as 1% in the pooled samples and eliminate SNPs that are not polymorphic in the test population. Pooling also provides high throughput comparisons between cases and controls to prioritize SNPs for subsequent individual genotyping. Efficiency gains can be made by using duplex assays and the use of MDA may reduce the effort required for pool construction.

Acknowledgments

This study was partly supported by grants to GWM from the National Health and Medical Research Council (339430 and 339446) and by the Cooperative Research Centre for the Discovery of Genes for Common Human Diseases established and supported by the Australian Government's Cooperative Research Centre's Program.

References

Bansal, A., van den Boom, D., Kammerer, S., Honisch, C., Adam, G., Cantor, C. R., Kleyn, P., & Braun, A. (2002). Association testing by DNA pooling: An effective initial screen. *Proceedings of the National Academy of Sciences USA, 99*, 16871–16874.

Braun, A., Little, D. P., & Koster, H. (1997). Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. *Clinical Chemistry, 43*, 1151–1158.

Bray, M. S., Boerwinkle, E., & Doris, P. A. (2001). High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: Practice, problems and promise. *Human Mutation, 17*, 296–304.

https://doi.org/10.1375/twin.8.4.353 Published online by Cambridge University Press

Buetow, K. H., Edmonson, M., MacDonald, R., Clifford, R., Yip, P., Kelley, J., Little, D. P., Strausberg, R., Koester, H., Cantor, C. R., & Braun, A. (2001). High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proceedings of the National Academy of Sciences USA, 98*, 581–584.

Cardon, L. R., & Bell, J. I. (2001). Association study designs for complex diseases. *Nature Reviews Genetics, 2*, 91–99.

Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S. F., Egholm, M., & Lasken, R. S. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences USA, 99*, 5261–5266.

Ding, C., & Cantor, C. R. (2003). A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. *Proceedings of the National Academy of Sciences USA, 100*, 3059–3064.

Higgins, G. S., Little, D. P., & Koster, H. (1997). Competitive oligonucleotide single-base extension combined with mass spectrometric detection for mutation screening. *Biotechniques, 23*, 710–714.

Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G., Frazer, K. A., & Cox, D. R. (2005). Whole-genome patterns of common DNA variation in three human populations. *Science, 307*, 1072–1079.

Hoogendoorn, B., Norton, N., Kirov, G., Williams, N., Hamshere, M. L., Spurlock, G., Austin, J., Stephens, M. K., Buckland, P. R., Owen, M. J., & O'Donovan, M. C. (2000). Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools. *Human Genetics, 107*, 488–493.

Hosono, S., Faruqi, A. F., Dean, F. B., Du, Y., Sun, Z., Wu, X., Du, J., Kingsmore, S. F., Egholm, M., & Lasken, R. S. (2003). Unbiased whole-genome amplification directly from clinical samples. *Genome Research, 13*, 954–964.

Jiang, R., Duan, J., Windemuth, A., Stephens, J. C., Judson, R., & Xu, C. (2003). Genome-wide evaluation of the public SNP databases. *Pharmacogenomics, 4*, 779–789.

Jurinke, C., Zollner, B., Feucht, H. H., van den Boom, D., Jacob, A., Polywka, S., Laufs, R., & Koster, H. (1998). Application of nested PCR and mass spectrometry for DNA-based virus detection: HBV-DNA detected in the majority of isolated anti-HBc positive sera. *Genetic Analysis, 14*, 97–102.

Lander, E. S. (1996). The new genomics: Global views of biology. *Science, 274*, 536–539.

Le Hellard, S., Ballereau, S. J., Visscher, P. M., Torrance, H. S., Pinson, J., Morris, S. W., Thomson, M. L., Semple, C. A., Muir, W. J., Blackwood, D. H., Porteous, D. J., & Evans, K. L. (2002). SNP genotyping on pooled DNAs: Comparison of genotyping technologies and a semi automated method for data storage and analysis. *Nucleic Acids Research, 30*, 74–83.

Little, D. P., Braun, A., Darnhofer-Demar, B., & Koster, H. (1997). Identification of apolipoprotein E polymorphisms using temperature cycled primer oligo base extension and mass spectrometry. *European Journal of Clinical Chemistry and Clinical Biochemistry, 35*, 545–548.

Little, D. P., Braun, A., O'Donnell, M. J., & Koster, H. (1997). Mass spectrometry from miniaturized arrays for full comparative DNA analysis. *Nature Medicine, 3*, 1413–1416.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research, 16*, 1215.

Mohlke, K. L., Erdos, M. R., Scott, L. J., Fingerlin, T. E., Jackson, A. U., Silander, K., Hollstein, P., Boehnke, M., & Collins, F. S. (2002). High-throughput screening for evidence of association by using mass spectrometry genotyping on DNA pools. *Proceedings of the National Academy of Sciences USA, 99*, 16928–16933.

Norton, N., Williams, N. M., Williams, H. J., Spurlock, G., Kirov, G., Morris, D. W., Hoogendoorn, B., Owen, M. J., & O'Donovan, M. C. (2002). Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. *Human Genetics, 110*, 471–478.

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science, 273*, 1516–1517.

Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature, 405*, 847–856.

Ross, P., Hall, L., & Haff, L. A. (2000). Quantitative approach to single-nucleotide polymorphism analysis using MALDI-TOF mass spectrometry. *Biotechniques, 29*, 620–629.

Schork, N. J., Cardon, L. R., & Xu, X. (1998). The future of genetic epidemiology. *Trends in Genetics, 14*, 266–272.

Sham, P., Bader, J. S., Craig, I., O'Donovan, M., & Owen, M. (2002). DNA Pooling: A tool for large-scale association studies. *Nature Reviews Genetics, 3*, 862–871.

Smith, D. J., & Lusis, A. J. (2002). The allelic structure of common disease. *Human Molecular Genetics, 11*, 2455–2461.

Tabor, H. K., Risch, N. J., & Myers, R. M. (2002). Opinion. Candidate-gene approaches for studying

https://doi.org/10.1375/twin.8.4.353 Published online by Cambridge University Press

complex genetic traits: Practical considerations. *Nature Reviews Genetics, 3*, 391–397.

Treloar, S., Hadfield, R., Montgomery, G., Lambert, A., Wicks, J., Barlow, D. H., O'Connor, D. T., Kennedy, S., & Group, I. E. S. (2002). The International Endogene Study: A collection of families for genetic research in endometriosis. *Fertility and Sterility, 78*, 679–685.

Visscher, P. M., & Le Hellard, S. (2003). Simple method to analyze SNP-based association studies using DNA pools. *Genetic Epidemiology, 24*, 291–296.

Werner, M., Sych, M., Herbon, N., Illig, T., Konig, I. R., & Wjst, M. (2002). Large-scale determination of SNP allele frequencies in DNA pools using MALDI-TOF mass spectrometry. *Human Mutation, 20*, 57–64.

https://doi.org/10.1375/twin.8.4.353 Published online by Cambridge University Press