

Chapter 4

Microarray-Based Genome-Wide Association Studies Using Pooled DNA

Szabolcs Szelinger, John V. Pearson, and David W. Craig

Abstract

Pooling genomic DNA samples within clinical classes of disease for use in whole-genome single nucleotide polymorphism (SNP) genotyping allows for rapid and inexpensive genome-wide association studies (GWAS). We describe here a general outline for combining hundreds of genomic DNA samples prior to genotyping on commercially available high-density SNP microarrays. The pool construction approach is universal, and independent of the SNP genotyping platform utilized, and therefore provides a quick, efficient, and low-cost alternative to interrogating thousands of individual samples on singular SNP microarrays. While the strategy for pooled DNA genotyping on SNP microarrays is straightforward, the success of such studies is critically dependent upon the accuracy of allelic frequency calculations, the ability to identify falsely positive results arising from assay variability, and the willingness to better resolve association signals through investigation of neighboring SNPs.

Key words: Genome-wide association study, Single nucleotide polymorphism, DNA pooling, Equimolar concentration, Allelic frequency, GenePool

1. Introduction

While 99.9% of the human genome sequence is identical in human beings, it is estimated that there are approximately ten million single nucleotide polymorphisms (SNPs) that, in combination, can distinguish among individuals (1–3). SNPs are nucleotide variants, for example, an adenosine to thymine (A/T) substitution, that occur throughout the human genome. These variants can contribute directly to disease predisposition by modifying a gene's function, or they can be used as genetic markers to detect nearby disease-causing mutations through association or family-based linkage studies. There are three general classes of SNPs: those with strong functional effects, which dramatically alter a

gene's behavior (classic mutations); those with more subtle functional effects that predispose to disease in concert with an individual's genetic background or environment (functional SNPs); and those which are completely silent with respect to function (nonfunctional SNPs). The immediate importance of genotyping a high-density panel of SNPs does not necessarily come from their functional relevance, but rather the proximity of the marker to the disease-causing variant or functional SNP (7). These types of studies, while long anticipated, are only recently possible for complex diseases (5, 6), and genotyping hundreds of individual DNA samples is a popular approach taken by researchers conducting GWAS. However, a major weakness of genome-wide genotyping approaches is the significant cost associated with purchase of the necessary reagents and microarrays, putting such studies beyond the reach of many research groups. We have successfully (7) developed an alternative approach of genotyping pooled instead of individual genomic DNA samples to reduce the overall cost of GWA studies. Several previous reports have investigated the feasibility of pooling on SNP genotyping microarrays or related technologies (8–11). With a few exceptions, these reports have focused on predicting allelic frequencies across thousands of SNPs rather than on the effectiveness of pooling in identifying the genetic basis of a specific complex disorder. For individual genotyping, there are a number of factors that influence the ability of GWA studies to detect genetic associations. These include, but are not limited to: (1) the allele frequency of the causal variant; (2) the odds ratio (OR) or genetic relative risk; (3) the linkage disequilibrium (LD) between the causal variant and the probed SNPs; (4) the number of individuals in each cohort; (5) the number of probed SNPs in LD with the causal variant; and (6) the analysis approach taken. Specific to pooling, there are additional factors that influence the ability to detect a true association. These additional factors include (7) the precision of allele frequency measurements made by the SNP genotyping microarray; (8) the accuracy of pool construction by pipetting; (9) the integrity of the pooled genomic DNA; (10) the number of individuals pooled or overall pooling strategy; and (11) the number of microarray technical replicates. Furthermore, population stratification and admixture can mask true associations in all studies. Beyond these additional factors, in pooling, one loses the ability to compare subphenotypes of pools, to directly measure genotype, and to detect gene–gene interactions. However, the most important factor in favor of pooling-based GWA studies is that this study design can be completed for thousands of dollars, whereas individual genotyping may require millions of dollars. There are numerous orphan diseases and many small populations which cannot realistically be studied using individual genotyping at this time, and a pooling-based GWA study presents an attractive, cost-effective alternative.

2. Materials

2.1. Quality Control and Quantification of Genomic DNA

2.1.1. Quality Control Using Gel Electrophoresis

1. Molecular biology grade water.
2. 50× Tris-acetate–EDTA buffer.
3. GelStar nucleic acid gel stain 10,000× (Lonza; Rockland, ME).
4. DNA ladder.
5. Ultrapure agarose.
6. 96-well Gel System (Thermo Fisher Scientific; Waltham, MA).
7. Microwave.
8. Magnetic stir plate and magnetic stir bars.
9. Dark Reader (Clare Chemical Research; Dolores, CO).

2.1.2. Quality Control and Quantification Using PicoGreen

1. Quant-iT PicoGreen dsDNA kit (Invitrogen; Carlsbad, CA).
2. Costar 96-well, black, flat bottom plates (Corning Life Sciences; Lowell, MA).
3. Molecular biology grade water.
4. Multi- and single-channel pipettes.
5. Aluminum foil.
6. Sterile glass bottle for 1× Tris–EDTA working solution.
7. Microcentrifuge.
8. Vortex.
9. Sterile tubes (1.5 mL).
10. Falcon tubes (50 mL, 15 mL).
11. Disposable reagent reservoirs.
12. Human genomic DNA Cat#:11 691 112 001 (Roche Applied Science; Indianapolis, IN).

2.2. Pool Construction

1. Single-channel pipettes (P1, P2, P10, P20).
2. Sterile collection tube (1.5–5.0 mL).
3. Ice bucket.
4. Parafilm.

2.3. Genotyping

Materials for pooled genotyping can vary and depend primarily on the infrastructure, funding, and choices of the researcher. In terms of this chapter, however, the array type is irrelevant, as the focus here is on a pooled genomic DNA technique that is independent of genotyping platform. For interested readers, specific protocols for some of the genotyping assays are described within other sections of this book. That said, the predominant array types include those manufactured by both

Illumina (i.e., the Human 610-Quad v1.0, Human 1M-Duo v3.0, Human CNV370-Quad v3.0) and Affymetrix (i.e., the Genome-Wide Human SNP Array 5.0 and Genome-Wide Human SNP Array 6.0).

2.4. Analysis of Pooled Genotype Data

GenePool software (freely available from <http://genepool.tgen.org>).

3. Methods

3.1. Quality Control and Quantification of Genomic DNA

3.1.1. Quality Control and Quantification Using Agarose Gel Electrophoresis

The visualization of DNA samples to be pooled on an agarose gel is not only an excellent way to control for the use of DNA samples of suboptimal quality, but it is also a great reference for use during troubleshooting procedures (see Notes 1 and 2). Agarose gel electrophoresis has long been associated with fluorometric assays that demanded the use of hazardous compounds such as ethidium bromide and an excitation wavelength in the ultraviolet range, both of which increase health risks to the researcher. However, novel, less toxic, fluorescent gel-staining products have become available, which are more sensitive for nucleic acid visualization and do not promote the health risks associated with ethidium bromide and ultraviolet radiation. Thus, we recommend fluorescent gel-staining products for DNA visualization during agarose gel electrophoresis.

1. Thaw the Gel Star stain at room temperature for 10–15 min (see Note 3).
2. Dilute the 50× TAE buffer to a 1× working concentration with molecular biology grade water.
3. Prepare a 1% agarose gel using 1× TAE (e.g., 2 g Ultrapure agarose for a 200 mL gel).
4. Heat mixture in microwave until agarose is dissolved.
5. Place agarose solution on stir plate and agitate gently until the temperature of the mixture reaches 55°C.
6. Add Gel Star stain (10,000× stock) to agarose solution to achieve a 1× concentration (see Note 4).
7. Mix stain and agarose on stir plate, then pour into gel tray and let sit at room temperature until solidified.
8. Place the tray containing the gel in the electrophoresis apparatus and cover with 1× TAE. Load the appropriate DNA ladder and DNA samples with running dye and separate fragments with a constant voltage.
9. Visualize the DNA bands using Darkreader's blue light between 400 and 500 nm.

*3.1.2. Quality Control
and Quantification Using
Quant-iT dsDNA PicoGreen*

1. Precise quantification of genomic DNA is critical for the success of a pooling approach to GWAS (see Note 5). This protocol assumes high-throughput sample quantification and the following description is for large-scale studies requiring the use of several 96-well Costar plates filled to capacity (see Note 6). However, the framework of the protocol is the same for scaled-down pooling studies as well. The most critical part of the assay is the creation of the standard curve. The reference human genomic DNA comes in a 200 ng/ μ L stock concentration; therefore, the DNA to be quantitated should be at a lower concentration. Prior to standard curve preparation this could be achieved by a quick assessment of the sample DNA concentration on a Nanodrop or a 260 absorbance reading, for example. DNA concentration should fall somewhere in the middle of the standard curve. After the approximate concentration of the DNA is determined, all samples to be pooled should be diluted with sterile water to an approximate concentration of 100 ng/ μ L.
2. Dilute the 20 $\times$ TE buffer, included in the Quant-iT kit, to a 1 $\times$ working concentration (see Note 7).
3. Create the eight-point standard curve by serial dilution of the stock Roche DNA (see Note 8). Use 1.5 mL sample tubes with 1 $\times$ TE buffer to a final concentration series of 0, 3.125, 6.25, 12.5, 25, 50, 100, and 200 ng/ μ L. Assay each point of the standard curve in triplicate and use the median concentration of each to generate the curve.
4. Label the Costar 96-well, black assay plate with sample/plate name.
5. Keep all samples to be assayed on ice throughout the entire procedure.
6. Add 99 μ L of 1 $\times$ TE to each well of the assay plate.
7. Add 1 μ L of sample DNA to each well of the assay plate in triplicate or quintuplicate in rows A–F, and 1 μ L of DNA standard to rows G–H (see Note 9). Cover plate with plate seal, vortex briefly, and spin at 1,000 rpm for 30 s to collect the TE and DNA to the bottom of the wells.
8. Dilute the PicoGreen reagent (150 μ L of stock reagent to 19.85 mL 1 $\times$ TE). Prepare dilution in a plastic Falcon tube and wrap in aluminum foil to prevent photodegradation. PicoGreen reagent has a light orange/magenta color.
9. Pour the diluted PicoGreen reagent into a large trough and add 100 μ L to each well of the assay plate containing the DNA using a multichannel pipette. Remove any bubbles from the bottom of the wells, as they can influence the fluorescence signal detected by the reader. Cover the plate with aluminum foil and let stand for 5 min at room temperature.

10. Read and record the fluorescence signal in plate reader. DNA-bound fluorescent dye is excited at 480 nm and intensity measured at 520 nm.
11. Calculate median DNA concentrations for each sample (see Note 10).

3.2. Pool Construction

1. Calculate the desired amount of genomic DNA to combine for each individual sample, and list in ascending order on a pooling sheet with the corresponding well number (see Note 11).
2. Load the samples to be pooled on a 96-well pooling plate corresponding to the pooling sheet.
3. Once this sheet is printed, place the DNA plate on ice and label a 1.5–5 mL, sterile collection tube with the desired pool ID, date, and DNA concentration.
4. Using a single-channel pipette, transfer the desired amount of DNA, per pooling sheet, into the collection tube. Once the sample is added to the collection tube, check off the corresponding number on the pooling sheet to avoid errors and track progress (see Notes 12, 13, 14).
5. Once all samples from the DNA plate are pooled, determine the concentration of each pool in triplicate using PicoGreen assay.

3.3. Genotyping

The study design in this chapter suggests the combined usage of both Illumina and Affymetrix SNP genotyping arrays. Both Illumina and Affymetrix assays are performed according to vendor protocols, described elsewhere this book.

3.4. Analysis of Pooled Genotype Data

We exclusively use the GenePool software package to detect shifts in relative allele frequency between pooled genomic DNA from cases and controls using Affymetrix or Illumina SNP-based genotyping microarrays. GenePool is written in C++ (gpextract) and C (gpanalyze). These programs can be run individually using command line Unix. GenePool can be downloaded from the GenePool Web site at <http://genepool.tgen.org/>. The software is currently provided as a precompiled binary for X86-Linux and as a source code. Manual pages for all executables are bundled in both source and binary distributions and are also available from the GenePool Web site in PDF and HTML formats for online viewing (7).

3.5. Gpextract

The gpextract program is the first of the three programs in the GenePool system. For Affymetrix analysis, gpextract uses Affymetrix's Fusion library to extract intensity values from the Affymetrix CEL files and write them out to a more compact customized binary file format that will be read by gpanalyze. The new files have the same name as the original CEL file but with the

string .gpb appended to each filename. For Illumina analysis, gpextract uses the raw .txt Illumina files to extract intensity values and writes them to a more compact, customized binary file format that will be read by gpanalyze. Note that for Illumina chips, gpextract may report extracting intensity values for more SNPs than is stated for a given chip since there are SNPs on the chip that are used only for quality control and testing purposes. In addition, each Illumina chip may have a different number of SNPs and intensity values since the distribution of beads on the chip is random.

3.6. Gpanalyze

The gpanalyze program takes the intensity values from the custom binary file (.gpb) and uses a variety of distance measures to assign a probability score to each SNP where the score indicates how unlikely the observed difference in allele frequency is between the hybridizations for the two DNA pools.

3.7. Gpcommand

The gpcommand program is the third of the three programs in the GenePool system. It is a perl script that helps users run basic pooling analyses by reading a configuration file and automatically invoking the other two GenePool programs gpextract and gpanalyze. The programs gpextract and gpanalyze are designed for direct use by researchers; however, each has a large number of command line options which can, initially, be difficult to understand and use correctly. gpcommand is a perl wrapper that will create and execute basic command lines that run gpextract and gpanalyze. All commands executed by gpcommand can optionally be echoed to a log file for later inspection.

3.8. Affymetrix 500K Arrays

For Affymetrix 500K arrays, the following protocol is recommended. Probe intensity data are read from CEL files and relative allele signal (RAS) values are calculated for each quartet. RAS values correspond to the ratio of the A probe to the sum of the A and B probes (where A is the major allele and B is the minor allele) and provide a quantitative index correlating to allele frequencies in pooled DNA. These values yield independent measures of different hybridization events and are consequently treated as individual data points. A silhouette statistic, which represents the mean of the distance of a point to all other points in its class (i.e., case pool) versus points in the other class (i.e., control pool), is used to rank all genotyped SNPs. Silhouette statistics range from 1, where complete separation between pools has been achieved, to -1, where it is not possible to distinguish between case and control pools. The calculation for a silhouette statistic is given as follows:

$$S = \frac{\sum_{i=1 \dots N} s(i)}{N}, \quad s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad b(i) = \min\{b(i)\}, \quad (1)$$

where the overall silhouette statistic, S , is the average of all of the individual silhouette values, $s(i)$, for each of the class comparisons, N refers to the number of replicate measures, I represents the replicate array within a class, $a(i)$ refers to the average Euclidian distance of RAS_a and RAS_s for replicate i of a class to all other replicates within its class, and $b(i)$ refers to the average distance of a replicate (i) to all replicates not within its class. This particular test-statistic has been heuristically shown to perform well at identifying SNPs with large allelic frequency differences for the Affymetrix 500K arrays only.

3.9. Illumina and Affymetrix 5.0/6.0

For the Illumina and Affymetrix arrays, we utilize a modified test of two proportions based on predicted allelic frequencies. Allelic frequencies are approximated using a k -correction method such that $f = A/(A + kB)$, where k is the correction factor for the ratio of the intensity of A and B probes. The pooling-test statistic, T , is given as follows:

$$T = \frac{|f_a - f_c|}{\sqrt{\bar{f}(1 - \bar{f}) (1 / 2n_a + 1 / 2n_c) + s_r^2 (1 / m_a + 1 / m_c)}}, \quad (2)$$

where f_a and f_b are the respective predicted allelic frequencies for the a and c cohorts, n_a and n_c are the number of individuals pooled, s_r^2 is the standard deviation for technical replicates, and m_a and m_c are the number of replicate arrays.

For analysis, one must specify the chip files (CEL for Affymetrix and .TXT for Illumina) that are in each of the respective cohorts. One must also specify the number of individuals within each cohort. All other values are computed during the analysis, and a ranked list of SNPs by the test-statistic is produced.

Depending on the study, one must then individually genotype SNPs to determine a p -value. P -values are not computed from the analysis of pooled data since any number of spurious environmental factors can throw off analysis substantially. In the end, pooling can be viewed as a screening tool, but does not replace individual genotyping.

3.10. Validation of Analysis Results

Various approaches exist for validating analysis results. Essentially, one individually genotypes SNPs identified during analysis. The types of methods include Illumina GoldenGate, Illumina iSelect, Illumina BeadArray/Veracode, Sequenom MassArray genotyping, Applied Biosystems TaqMan, direct sequencing, and restriction-fragment length polymorphism (RFLP) analysis. Most of these methods are discussed elsewhere in this text. Due to the subjective nature for a researcher's choice in validating their results and also because of space constraints, we did not describe

the methods for individual genotyping here. However, we do encourage researchers to validate the analysis results, as the pooling-based GWAS should only be viewed as a screening tool for the SNPs showing the greatest differences in allele frequency between pools. When reporting results, only p -values from individual genotyping should be considered.

4. Notes

1. Sample DNA amplified through the use of available whole-genome amplification technologies is not suggested due to uneven amplification.
2. Like most molecular biology applications, the quality of the template DNA greatly determines the success of the study. Therefore, great care must be taken to ensure that the DNA used in pooling-based genotyping is of high quality and void of protein and other contaminants. We suggest that DNA samples are plated out on a 96-well microplate prior to study for easy handling. The layout of this “DNA plate” should be used for subsequent procedures.
3. Gel Star does not only stain dsDNA but ssDNA and RNA as well. However, it is several orders of magnitude more sensitive to dsDNA than RNA or ssDNA, which are common parts of purified DNA solutions. Other stains that are more specific to dsDNA are recommended for DNA quantification.
4. A 1× stain concentration in the agarose solution is recommended by the manufacturer, e.g., 5 μ L of Gel Star stain is added to every 50 mL of agarose solution. However, we achieved comparable results using only 0.2× of stain concentration in the gel, e.g., 2 μ L of Gel Star stain to 100 mL of agarose solution.
5. Precise quantification of genomic DNA is critical for the success of a pooling approach to GWAS. Because pooling-based genotyping measures the allelic frequency of a genetic variant using pools of genomic DNA partitioned into risk and non-risk groups, inaccuracies in the estimation of DNA concentration could potentially mask a predicted allele frequency difference between affected and unaffected pools. The Quant-iT kit is a rapid, accurate, and cost-effective method when working with dozens to hundreds of DNA samples simultaneously. Not only does this method enable high-throughput application, but PicoGreen results are also less influenced by common sample contaminants such as residual proteins, salts, or RNA. However, it should be noted that the PicoGreen reagent is sensitive to photodegradation and thus,

should be kept in a dark place, preferably at 4°C. Further, while performing the quantification assay, all prepared reagents should be kept in a plastic container, as Picogreen reagent adheres easily to glass surfaces. The method used by the authors follows the basic outline of the manufacturer's protocol, but adjusted to accommodate high-throughput requirements. Human genomic DNA standard is used instead of the calf thymus DNA (supplied with the Quant-iT kit) to more closely emulate the properties of the sample DNA. At the very least, sample DNA concentration should be measured in triplicate, but for higher accuracy, we recommend five replicate measurements for each sample to be pooled. The increased replicate count will inherently confirm some outlier measurements that would preferentially increase or decrease the actual concentration. Thus, more replicates with the use of median instead of mean will likely give the most accurate concentration estimate.

6. Pooling studies usually require dozens to hundreds of samples to be combined; therefore working with DNA samples in a 96-well microplate format greatly increases speed, simplicity, and precision. To reduce error during sample handling, the same sample layout should be used for the entire quantification and pooling process.
7. If large numbers of samples are to be assayed, the 1× TE buffer can be made up in a large batch and stored in a labeled glass container at room temperature for an extended period of time.
8. The Roche DNA should never be kept at −20°C. We observed consistently increased variance of R^2 values below 0.8 for DNA standard made out of previously frozen Roche reference DNA.
9. We recommend that rows G and H be dedicated for the diluted Roche DNA serial diluted standard. Thus, once the 99 µL of 1× TE is added to these wells, the standards are loaded in triplicate. The 0 ng/µL standard taking positions G1, G2, and G3, the 3.125 ng/µL standard taking up the positions G4, G5, G6 wells, and so on, until the wells H10, H11, and H12 contain the most concentrated standard 200 ng/µL. Just like the DNA samples of interest, these standards (1 µL) are added to each reaction well. For clarity and straightforwardness, a standard plate layout can be created and used for all subsequent assays. A standardized layout will also require only a single protocol in a fluorescent dye reader like Bio-Tek KC4.
10. Samples with greater than 2 standard deviations from the median volume to be pooled for each sample are not to be included in the final pool. Other samples are pooled at equimolar concentrations, ensuring the same number of DNA molecules represented in the combined DNA pool.

If concentration measurement of sample to be pooled is inaccurate or not equimolarly pooled, and pipetting errors occur, there is likely to be bias toward the allelic signal that is generated by over-, or under-represented genomic loci.

11. By pooling, one loses the power to detect associations, thus, we suggest study designs with multiple subpools containing an equal number of samples at equimolar concentration. The creation of subpools does not increase the power but allows for the reduction in pooling variance by pipetting error during pool construction. To increase resolution, we suggest the application of as many SNP platforms as possible with each subpool run in triplicate technical replicates. The redundancy between the SNP platforms allows for the reduction of measurement noise on highly correlated neighboring SNPs in LD and insures that there are very few gaps in overall genome-wide SNP coverage. The pool concentration is based on the obtained median concentrations of individual samples by the PicoGreen assay. Based on the median concentration of each individual sample, pooling volumes of each DNA sample is calculated; so all samples will contribute the same amount of DNA to each pool. At this point, the number of samples for each subpool is determined. The pool concentration is calculated to be approximately 10 $\times$ greater than the template concentration needed for the genotyping assay. This stock pool can be later diluted to the working concentration as desired by the various genotyping assays. If the concentration of DNA sample by Picogreen assay is 10 ng/ μ L, for example, and the desired final pool concentration is 100 ng/ μ L, the sampled DNA volume is $100/10 = 10 \mu$ L as 10 μ L of sample contains 100 ng of DNA to be pooled. The above example can be automated in a spreadsheet for all samples.
12. No less than 1 μ L of the sample in question is added to the pool. Volumes below 1 μ L increase pipetting error significantly.
13. For clarity and reducing error, the use of an unopened box of pipette tips is recommended. Pipette the sample DNA from the well A1 of the pooling plate using the pipette tip from the corresponding location on the pipette box. This will ensure that a sample is not pooled more than once and helps keep track of progress.
14. The DNA samples on the pooling sheet should be sorted from smallest to largest volumes so changing volumes on the pipette is uncomplicated and straightforward. Use of a spreadsheet with pooling volumes listed in increasing order will prevent fatigue in hands caused by constant adjustment of pipette to desired volume. If the smallest volume is pooled first, then simply upward adjustment of pipette dial will be needed until the last sample is added to the pool.

References

1. Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299–309
2. Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA (2003) Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat Genet* 33:518–521
3. Kruglyak L, Nickerson DA (2001) Variation is the spice of life. *Nat Genet* 27:234–236
4. Schaid DJ, Guenther JC, Christensen GB, Hebbbring S, Rosenow C, Hilker CA, McDonnell SK, Cunningham JM, Slager SL, Blute ML, Thibodeau SN (2004) Comparison of microsatellites versus single-nucleotide polymorphisms in a genome linkage screen for prostate cancer-susceptibility loci. *Am J Hum Genet* 75(6):948–965
5. Botstein D, Risch N (2003) *Nat Genet* 33(Suppl):228–237
6. Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273:1516–1517
7. Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, Homer N, Brun M, Szelinger S, Coon KD, Zismann VL, Webster JA, Beach T, Sando SB, Aasly JO, Heun R, Jessen F, Kolsch H, Tsolaki M, Daniilidou M, Reiman EM, Papassotiropoulos A, Hutton ML, Stephan DA, Craig DW (2007) Identification of the genetic basis for complex disorders by use of pooling-based genome-wide single-nucleotide-polymorphism association studies. *Am J Hum Genet* 80:126–139
8. Bansal A et al (2002) Association testing by DNA pooling: an effective initial screen. *Proc Natl Acad Sci USA* 99:16871–16874
9. Meaburn E, Butcher LM, Schalkwyk LC, Plomin R (2006) Genotyping pooled DNA using 100K SNP microarrays: a step towards genome-wide association scans. *Nucleic Acids Res* 34:e27
10. Hinds DA et al (2004) Application of pooled genotyping to scan candidate regions for association with HDL cholesterol levels. *Hum Genomics* 1:421–434
11. Barratt BJ et al (2002) Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design. *Ann Hum Genet* 66:393–405



<http://www.springer.com/978-1-61737-953-6>

Disease Gene Identification

Methods and Protocols

DiStefano, J.K. (Ed.)

2011, XII, 312 p., Hardcover

ISBN: 978-1-61737-953-6

A product of Humana Press