Methods 52 (2010) 255–258

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

DNA methylation profiling using bisulfite-based epityping of pooled genomic DNA

Sophia J. Docherty *, Oliver S.P. Davis, Claire M.A. Haworth, Robert Plomin, Jonathan Mill

King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK

ARTICLE INFO

Article history:
Available online 3 July 2010

Keywords:
Epigenetics
DNA pooling
Sodium bisulfite
DNA methylation
High-throughput

ABSTRACT

DNA methylation plays a vital role in normal cellular function, with aberrant methylation signatures being implicated in a growing number of human pathologies and complex human traits. Methods based on the modification of genomic DNA with sodium bisulfite are considered the 'gold-standard' for DNA methylation profiling on genomic DNA; however they require large amounts of DNA and may be prohibitively expensive when used on the large sample sizes necessary to detect small effects. DNA pooling approaches are already widely used in large-scale studies of DNA sequence and gene expression. In this paper, we describe the application of this economical DNA pooling technique to the study of DNA methylation profiles. This method generates accurate quantitative assessments of group DNA methylation averages, reducing the time, cost and amount of DNA starting material required for large-scale epigenetic investigation of disease phenotypes.

1. Introduction

1.1. Growing interest in epigenetic research

Epigenetics refers to the reversible regulation of various genomic functions mediated through partially stable modifications of DNA and chromatin histones. Epigenetic processes are essential for normal cellular development and differentiation, and allow the regulation of gene function through non-mutagenic mechanisms. Of particular interest is the phenomenon of cytosine methylation, occurring at position 5 of the cytosine pyrimidine ring in CpG dinucleotides. This process is intrinsically linked to the regulation of gene expression, with many genes demonstrating an inverse correlation between the degree of DNA methylation and the level of expression [1]. The methylation of these CpG sites, over-represented in CpG-islands in the promoter regulatory regions of many genes, disrupts the binding of transcription factors and attracts methyl-binding proteins that are associated with gene silencing and chromatin compaction. DNA methylation plays a vital role in normal cellular function, and aberrant methylation signatures have been implicated in a growing number of human pathologies [2,3] including cancer [4], imprinting disorders [5], and even complex neuropsychiatric phenotypes such as schizophrenia and bipolar disorder [6].

The 'gold-standard' method for mapping methylated cytosines is via the treatment of genomic DNA with sodium bisulfite; this process converts unmethylated cytosines to uracils (and subsequently, via PCR, to thymidines), while methylated cytosines are resistant to bisulfite and remain unchanged [7]. After sodium bisulfite treatment, DNA regions of interest are amplified and interrogated to identify C → T transitions or stable C positions, respectively corresponding to unmethylated and methylated cytosines in the native DNA. Numerous methods of analyzing bisulfite-modified DNA have been described [8,9], including the use of next-generation deep-sequencing methodologies to enable the highly-parallel analysis of bisulfite-treated samples [10–12].

1.2. Financial obstacles to the study of DNA methylation

Though deep-sequencing of large samples is currently economically infeasible for most researchers, a more accessible method which employs base-specific cleavage followed by MALDI-TOF mass spectrometry, can generate a quantitative estimate of the proportion of methylated DNA at a specific CpG site in a given sample [13,9]. Such highly-quantitative DNA methylation analysis is clearly vital to our understanding of gene function and the role of epigenetic dysfunction in disease, but wisdom gained following recent large-scale genetic association studies suggests that extremely large sample sizes may be crucial in detecting the small effects expected in the highly complex disorders that contribute most to the global burden of disease [14]. The expense of such large-scale research remains prohibitive to many researchers, and this economic obstacle is bolstered further by the relatively large quantities of DNA required for bisulfite treatment, especially

* Corresponding author. Address: Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, UK. Fax: +44 (0)207 848 0866.

E-mail addresses: Sophia.docherty@iop.kcl.ac.uk (S.J. Docherty), Oliver.davis@iop.kcl.ac.uk (O.S.P. Davis), Claire.haworth@iop.kcl.ac.uk (C.M.A. Haworth), r.plomin@iop.kcl.ac.uk (R. Plomin), j.mill@iop.kcl.ac.uk (J. Mill).

1046-2023/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2010.06.017

if multi-locus or whole-genome approaches are to be utilized, and by the fact that quantitative DNA methylation assessment, unlike genotypic assessment, requires technical replicates to ensure accuracy. Whilst the systematic assessment of DNA methylation has the potential to revolutionize our knowledge about the etiology of many complex disorders, current methods remain unsuitable for profiling the large sample cohorts likely to be required to detect pathogenic epimutations, especially for complex disorders or where multiple tissue-types need to be assessed.

1.3. The benefits of DNA pooling

Validated pooling techniques are widely employed to increase throughput in studies of DNA sequence variation [15,16] and gene expression [17], and have allowed researchers to assess samples of sizes which would otherwise be economically infeasible. Rather than generating individual results and averaging them within a group, such approaches combine the DNA of different individuals to generate direct estimates of their average result. These estimates can then be used to compare 'case' and 'control' groups, or groups from the 'high' and 'low' extremes of quantitatively-assessed traits. In studies of DNA sequence variation, screens of DNA pools across thousands of loci have been used to identify regions of the genome for further study via individual genotyping, yielding promising results [18–20]. Conversely, some screens have reported no group differences, indicating that the time and expense of further investigation at the individual-sample level may be unjustified [21]. Though individual genotype and haplotype data is lost, which may be particularly important if there are as yet unknown etiologically-relevant subgroups within cases of interest, the economic benefits of DNA pooling outweigh these disadvantages for many researchers.

Here we describe a high-throughput DNA pooling method, validated in 2009 [22], which uses bisulfite treatment followed by MALDI-TOF mass spectrometry to quantitatively assess group DNA methylation averages.

2. Methods

2.1. DNA pools

2.1.1. Pool design

In other pooling approaches to the study of DNA methylation, DNA pools have been created subsequent to bisulfite treatment [23]. However, as this approach could potentially be affected by differential bisulfite conversion biases and requires relatively large amounts of starting material from each sample, the current method involves pool construction prior to bisulfite treatment. A crucial consideration during pool design is the possibility of subgroups within cases of interest. When unknown etiological subgroups are present within a population, as has been shown through differing DNA methylation levels in many cancers, pooling strategies are at a clear disadvantage [24,25]. One option would be to include individuals in any known potentially important subgroups (e.g. grouping by age-of-onset, severity, disease-progression, chronicity, etc.) together in the same pool to allow for comparisons between these subsets of an overall 'case' group.

Pool number and size can also be important factors, and the sensible design of a DNA pooling study is critical to its cost-effectiveness. The cheapest design would involve the creation of a single large DNA pool for each group being compared. However, the creation of a number of pools containing fewer individuals allows for the inclusion of within-group variation in any analyses. Furthermore, the inclusion of fewer individuals into any one pool ensures that each individual will contribute a sufficient number of DNA molecules to allow for accurate analysis of the group. This may be especially important in the generation of pools intended for DNA methylation analysis, as DNA methylation levels are quantitatively assessed across a number of DNA molecules. On the other hand, the creation of numerous pools, each comprising only a few samples will result in all of the lost accuracy associated with pooling designs, but with little of the economic benefits.

Our experience has shown that in order to achieve a reasonable compromise between the amount of information lost in the pooling process and the time and expense saved, studies should include a few pools, each containing a relatively large number of individuals [17,26]. In the original validation of this method, pools containing the DNA of between 29 and 89 individuals were tested and shown to perform to the same high standard [22]. As well as the use of biological replicates, in the form of multiple DNA pools, the use of technical replicates is advised. Ideally this technical replication could commence at a number of stages of the protocol, to account for all of the possible variation involved. Just as with individual samples, bisulfite conversion, PCR and MALDI-TOF stages can be replicated on the same pools to average out errors. In addition, technical replication of pool creation is advised to account for any inconsistencies in the amounts of each sample added.

2.1.2. Sample preparation

The careful and precise construction of DNA pools, in which all individuals are equally represented, is essential to the accurate assessment of group averages in any pooling study. The quality of all DNA samples must be closely matched, as the inclusion of lower quality samples, likely to perform poorly, may lead to their under-representation in the final average result. It is therefore preferable that any single DNA pool only contain DNA samples extracted from the same tissue, using the same method. Additionally, it is advisable that all samples are tested for degradation, for example via agarose gel electrophoresis, and excluded if significant degradation is found. Where DNA quantities are extremely limited, samples may need to be tested sparingly. In our experience only ~10–20 ng of each DNA sample needs to be loaded onto a gel.

Each sample to be included must be accurately quantified. As this is incredibly important to the generation of equally-represented DNA pools, unless samples are of an extremely high purity, ultraviolet (UV) light spectroscopy – even using Nanodrop technology – will not be sufficiently accurate. We therefore recommend quantification using fluorimetry, employing a DNA-specific dye such as PicoGreen® dsDNA quantitation reagent (Cambridge Bioscience, UK). Prior to quantification, samples should be shaken overnight and inverted to ensure a homogeneous DNA concentration. In order to generate reliable quantification estimates, each sample should be quantified in triplicate. If the range of these three readings is greater than five percent of their average, we recommend re-quantification until the range is smaller.

2.1.3. DNA pool construction

Once quantified, all samples should be diluted to the same working concentration. This concentration may vary depending on the intended subsequent treatment of the pool. For example, certain commercially available bisulfite treatment kits are designed to process only a small volume of sample. Such specifications should be checked before constructing a pool too dilute for purpose. DNA quality is another consideration. When using lower quality DNA (e.g. DNA extracted from buccal swabs as opposed to blood), we would recommend bisulfite treating larger amounts of DNA, and (again with the specific method of bisulfite treatment's demands in mind) this may mean constructing more highly concentrated DNA pools. After the diluted samples have been left to diffuse overnight, and have again been shaken and inverted to ensure homogeneity, equal volumes – which will now also be equal amounts – of each sample can be combined to create a DNA pool. Pool construction should be undertaken with great care and precision, preferably by only one individual using the same pipette, to minimize any inconsistency. One may assess the accuracy of pool construction by genotyping SNPs in the DNA pools, and comparing the actual SNP allele frequencies within pools – based upon individual genotyping data – to those estimated from the DNA pools (see [27] for an example of this).

2.2. DNA methylation analysis

2.2.1. Assay design

Assays can be designed for target regions using the online Sequenom EpiDesigner software (www.epidesigner.com). Additional assay design assistance can be found in the MassArray R package [28], which also contains a number of useful tools for analyzing Sequenom output files.

2.2.2. Sodium bisulfite treatment

The protocol for the DNA methylation analysis of DNA pools is identical to that of individual samples. Any study should therefore include positive and negative controls. As well as a negative template control of water to detect contamination, positive controls of fully methylated and fully unmethylated DNA – available for example in the Human Methylated and Non-methylated DNA Set (Zymo Research, CA, USA) – should be used to check the validity of the assay. DNA pools and all controls should be bisulfite treated. We recommend using the EZ-96 DNA Methylation Kit (Zymo Research, CA, USA) following the manufacturers' standard protocol. We also recommend the use of Hot Star *Taq* DNA polymerase (Qiagen, UK) in bisulfite-PCR amplification.

2.2.3. Quantification of DNA pool methylation levels

DNA methylation analysis is conducted following bisulfite-PCR amplification using the Sequenom EpiTYPER system (Sequenom Inc., CA, USA) as described previously [29]. As with PCR amplification of untreated genomic DNA, some optimization of PCRs may be necessary to ensure specificity. Bisulfite treatment converts unmethylated cytosines to uracils, while methylated cytosines remain unchanged. These sequence changes are preserved during subsequent PCR, with conversion to thymidine at unmethylated (but not methylated) cytosine positions. The Sequenom EpiTYPER technique involves in vitro transcription of the amplified sequence, followed by enzymatic base-specific cleavage of the resulting RNA transcript. The exact weight of the fragments produced will depend upon bisulfite-treatment induced variations in the DNA sequence. MALDI-TOF mass spectrometry is used to assess the size ratio of the cleaved products, providing quantitative methylation estimates for CpG sites within a target region [13]. As with the analysis of individual samples, the running of technical replicates is advised to obtain the most reliable DNA methylation data. We highly recommend replication of the bisulfite treatment reaction to account for the variability introduced by differential conversion at this stage [30].

3. Results and discussion

This method was validated in 2009 [22] using 89 high-quality Centre d'Etude du Polymorphisme Humain (CEPH) genomic DNA samples extracted from transformed lymphoblastoid cell lines (Coriell Institute for Medical Research, NJ, USA). Four independent pools were formed from the DNA of CEPH: (1) 'Mothers' ($N = 29$), (2) 'Fathers' ($N = 30$), (3) 'Offspring' ($N = 30$), and (4) the entire sample ($N = 89$). The four pools were processed alongside the 89 individual samples, with technical replication conducted from

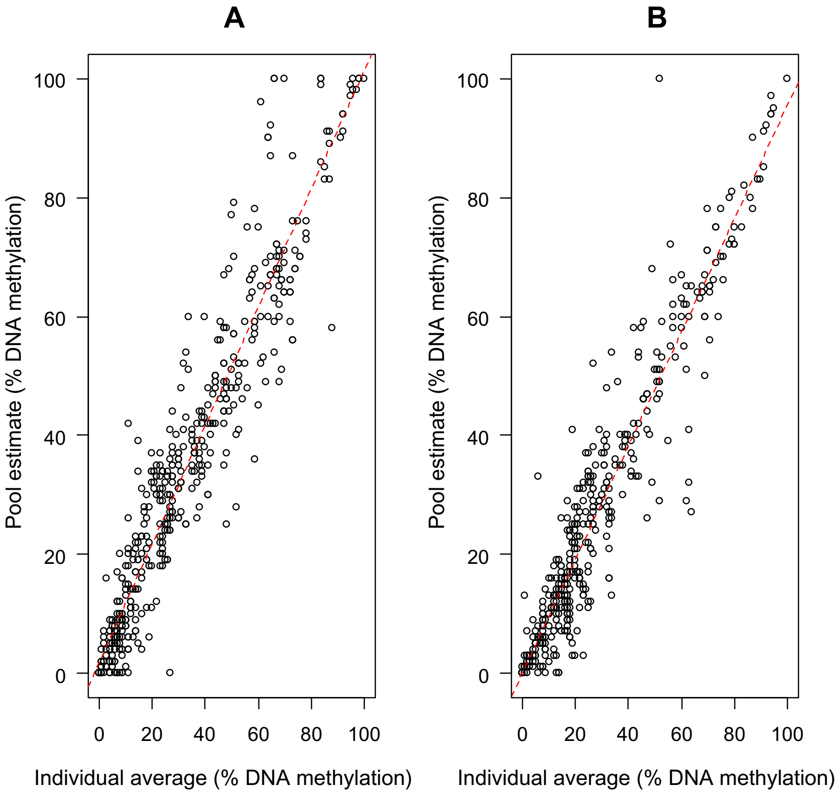

Fig. 1. Taken from [22] DNA methylation estimates across 205 CpG sites obtained from pooled DNA samples are highly correlated with actual average DNA values obtained from individual DNA samples in both (A) the first and (B) the second technical replicates of the validation experiment.

the bisulfite treatment stage. 205 CpG sites (133 CpG units) spread across 9 genomic regions were assessed. DNA methylation values obtained from individual samples were averaged across subjects within each pool, and these averages were compared to group DNA methylation estimates generated from the DNA pools. The overall correlation across all CpG sites assessed was 0.95 (95% bootstrapped confidence intervals: 0.94–0.96) in the first replicate (see Fig. 1A) and 0.95 (95% bootstrapped confidence intervals: 0.93–0.96) in the second replicate (see Fig. 1B), with an overall correlation of 0.95 (95% bootstrapped confidence intervals: 0.94–0.96) across the averaged data from both replicates. This correlation is comparable to the correlation of 0.95 between technical replicates i.e. the results gathered from the same individual samples, averaged across individuals within each pool, in the first and second replicates. For further details and results from the validation study see Docherty et al. [22]. This method provides an economical and valid means to accurately estimate group DNA methylation averages.

4. Concluding remarks

We have described the application of the Sequenom EpiTYPER system to the analysis of pooled DNA, for estimating average DNA methylation levels within a group. This method can be readily used to detect group differences in the study of a wide range of disease phenotypes. As it reduces the time, cost and amount of DNA starting material required, such an approach may be especially useful to researchers with limited funds and DNA stocks. In large-scale studies involving multiple candidate regions, this economical method will also prove valuable in highlighting those regions of the genome which warrant further study at the individual-sample level.

References

[1] R. Jaenisch, A. Bird, Nature Genetics 33 (2003) 245–254.
[2] E. Hatchwell, J.M. Greally, Trends in Genetics 23 (2007) 588–595.
[3] K.D. Robertson, A.P. Wolffe, Nature Reviews Genetics 1 (2000) 11–19.
[4] P.A. Jones, S.B. Baylin, Cell 128 (2007) 683–692.
[5] A.P. Feinberg, Nature-London 447 (2007) 433.
[6] J. Mill, T. Tang, Z. Kaminsky, T. Khare, S. Yazdanpanah, L. Bouchard, P. Jia, A. Assadzadeh, J. Flanagan, A. Schumacher, The American Journal of Human Genetics 82 (2008) 696–711.
[7] S.J. Clark, J. Harrison, C.L. Paul, M. Frommer, Nucleic Acids Research 22 (1994) 2990.
[8] S.J. Clark, A. Statham, C. Stirzaker, P.L. Molloy, M. Frommer, Nature Protocols 1 (2006) 2353–2364.
[9] P.W. Laird, Nature Reviews Genetics 11 (2010) 191–203.
[10] K.H. Taylor, R.S. Kramer, J.W. Davis, J. Guo, D.J. Duff, D. Xu, C.W. Caldwell, H. Shi, Cancer Research 67 (2007) 8511.
[11] R. Lister, J.R. Ecker, Genome Research 19 (2009) 959.
[12] A. Meissner, T.S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B.E. Bernstein, C. Nusbaum, D.B. Jaffe, Nature 454 (2008) 766–770.
[13] M. Ehrich, M.R. Nelson, P. Stanssens, M. Zabeau, T. Liloglou, G. Xinarianos, C.R. Cantor, J.K. Field, D. van den Boom, Proceedings of the National Academy of Sciences of the United States of America 102 (2005) 15785–15790.
[14] Wellcome Trust Case Control Consortium, Nature 447 (2007) 661–678.
[15] S.J. Docherty, L.M. Butcher, L. Schalkwyk, R. Plomin, BMC Genomics 8 (2007) 214.
[16] G. Kirov, I. Nikolov, L. Georgieva, V. Moskvina, M.J. Owen, M.C. O'Donovan, BMC Genomics 7 (2006) 27.
[17] C. Kendziorski, R.A. Irizarry, K.S. Chen, J.D. Haag, M.N. Gould, Proceedings of the National Academy of Sciences of the United States of America 102 (2005) 4252–4257.
[18] S.J. Docherty, O.S.P. Davis, Y. Kovas, E.L. Meaburn, P.S. Dale, S.A. Petrill, L.C. Schalkwyk, R. Plomin, Genes, Brain, and Behavior 9 (2010) 234–247.
[19] L.M. Butcher, O.S.P. Davis, I.W. Craig, R. Plomin, Genes, Brain, and Behavior 7 (2008) 435–446.
[20] E.L. Meaburn, N. Harlaar, I.W. Craig, L.C. Schalkwyk, R. Plomin, Molecular Psychiatry 13 (2007) 729–740.
[21] L.M. Butcher, R. Plomin, Behavior Genetics 38 (2008) 361–371.
[22] S.J. Docherty, O.S.P. Davis, C.M.A. Haworth, R. Plomin, J. Mill, Epigenetics & Chromatin 2 (2009) 3.
[23] E. Dejeux, V. Audard, C. Cavard, I.G. Gut, B. Terris, J. Tost, Journal of Molecular Diagnostics 9 (2007) 510.
[24] K.J. Kron, L. Liu, V.V. Pethe, N. Demetrashvili, M.E. Nesbitt, J. Trachtenberg, H. Ozcelik, N.E. Fleshner, L. Briollais, T.H. van der Kwast, Laboratory Investigation 90 (2010) 1060–1067.
[25] P.W. Ang, M. Loh, N. Liem, P.L. Lim, F. Grieu, A. Vaithilingam, C. Platell, W.P. Yong, B. Iacopetta, R. Soong, BMC Cancer 10 (2010) 227.
[26] P. Sham, J.S. Bader, I. Craig, M. O'Donovan, M. Owen, Nature Review Genetics 3 (2002) 862–871.
[27] E. Meaburn, L.M. Butcher, L.C. Schalkwyk, R. Plomin, Nucleic Acids Research 34 (2006) e27.
[28] R.F. Thompson, M. Suzuki, K.W. Lau, J.M. Greally, Bioinformatics 25 (2009) 2164.
[29] M.W. Coolen, A.L. Statham, M. Gardiner-Garden, S.J. Clark, Nucleic Acids Research 35 (2007) e119.
[30] M. Ehrich, S. Zoll, S. Sur, D. van den Boom, Nucleic Acids Research 35 (2007) e29.