CLINICAL EPIGENETICS

METHODOLOGY

Open Access

CrossMark

DNA methylation levels are highly correlated between pooled samples and averaged values when analysed using the Infinium HumanMethylation450 BeadChip array

Cristina Gallego-Fabrega[1,2], Caty Carrera[3], Elena Muiño[1], Joan Montaner[3], Jurek Krupinski[4,5], Israel Fernandez-Cadenas[1*] and On behalf of Spanish Stroke Genetics Consortium

Abstract

Background: DNA methylation is a heritable and stable epigenetic mark implicated in complex human traits. Epigenome-wide association studies (EWAS) using array-based technology are becoming widely used to identify differentially methylated sites associated with complex diseases. EWAS studies require large sample sizes to detect small effects, which increases project costs. In the present study we propose to pool DNA samples in methylation array studies as an affordable and accurate alternative to individual samples studies, in order to reduce economic costs or when low amounts of DNA are available. For this study, 20 individual DNA samples and 4 pooled DNA samples were analysed using the Illumina Infinium HumanMethylation450 BeadChip array to evaluate the efficiency of the pooling approach in EWAS studies. Statistical power calculations were also performed to discover the minimum sample size needed for the pooling strategy in EWAS.

Results: A total of 485,577 CpG sites across the whole genome were assessed. Comparison of methylation levels of all CpG sites between individual samples and their related pooled samples revealed highly significant correlations (rho > 0.99, p-val < 10^{-16}). These results remained similar when assessing the 101 most differentially methylated CpG sites (rho > 0.98, p-val < 10^{-16}). Also, it was calculated that $n = 43$ is the minimum sample size required to achieve a 95 % statistical power and a 10^{-06} significance level in EWAS, when using a DNA pool strategy.

Conclusions: DNA pooling strategies seems to accurately provide estimations of averaged DNA methylation state using array based EWAS studies. This type of approach can be applied to the assessment of disease phenotypes, reducing the amount of DNA required and the cost of large-scale epigenetic analyses.

Background

Epigenetics refers to the stable, heritable and reversible modifications in DNA expression associated with transcriptional regulation without alterations in the nucleotide sequence [1]. Epigenetic processes such as DNA methylation (DNAm), histone acetylation/deacetylation, non-coding mRNA expression and chromatin conformational changes [2] are essential for normal cellular development and differentiation. They have also been

linked to some monogenic and complex human diseases [3, 4]. Nowadays DNA methylation is one of the most studied epigenetic modifications [5, 6] and alterations in methylation have been linked with some disease processes such as different types of cancer [4, 7, 8], as well as with aging and exposure to tobacco smoke [9–12].

Some of the most important technologies used to detect DNA methylation are: deep sequencing, high-throughput deep sequencing and array-based genome-wide studies such as Epigenome Wide Association (EWAS) [13].

In the "omics" era, Genome-wide Association Studies (GWAS) have been widely used to discover the genetic polymorphisms associated with human diseases. These

* Correspondence: israelcadenas@yahoo.es
[1]Stroke pharmacogenomics and genetics, Fundació Docència i Recerca Mutua Terrassa, Hospital Universitari Mútua de Terrassa, C/ Sant Antoni 19, 08221 Terrassa, Barcelona, Spain
Full list of author information is available at the end of the article

Gallego-Fabrega *et al. Clinical Epigenetics* (2015) 7:78

Page 2 of 8

studies have been more successful in finding genes associated with complex diseases, compared to classical candidate genes studies. However, GWAS needs higher sample sizes and specific arrays that increase project costs. Several papers have observed that the use of pooling strategies decreases the cost of GWAS, while providing similar results to individual sample analysis [14].

Pearson and colleagues reported that pooling-based GWAS was theoretically effective in identifying genetic associations in different types of disease [15]. Applying these methods to experimental case–control data, they also demonstrated the successful identification of previously published susceptible loci for a rare monogenic disease, a rare complex disease and a common complex disease. In addition, Gaj et al. confirmed previously reported loci for colorectal cancer and prostate cancer in a Polish population, with a pooled-based strategy using GWAS [16].

Epigenome-wide association studies (EWAS) use the same strategy as GWAS, but for epigenetics. EWAS use array-based genotyping technology to detect the methylation levels at CpG sites across the genome. EWAS of human diseases are becoming increasingly common [4, 7, 17, 18]. Like GWAS, the EWAS are hypothesis-free approaches to finding differentially methylated sites instead of different allele frequencies. Nevertheless, pooled DNA strategies might be an affordable alternative that reduces study costs in array-based EWAS.

No current studies have analysed the accuracy of DNA pooling strategies in array-based EWAS. Our aim is thus to analyse the pooling strategies in EWAS studies in order to determine the effectiveness of these approaches in studying DNA methylation patterns in human samples.

In the present study, data from 20 individual DNA samples and 4 pooled DNA samples, analysed with the Illumina Infinium HumanMethylation450 BeadChip, were used to estimate the feasibility of the pooling approach, comparing the results of the individual samples to the results of the DNA pools of the same samples.

Results and discussion

Quality control

A total of 485,577 CpG sites across the whole genome were assessed using the Illumnia HumanMethylation450 BeadChip, in 20 individual samples and 4 DNA pools. First, the distribution of methylation level was evaluated for all samples, with DNA pools showing the same behaviour as individual samples (Fig. 1). Before the quantile normalization, 33,301 CpG sites and no samples were removed due to QC issues.

Correlations

Data revealed highly significant correlations (p-value $< 10^{-16}$, Spearman's test) after comparing the data generated from the pooled DNA samples with the averaged results of the individual samples. Group A and group B samples were studied separately with their respective pools, the obtained correlations were rho = 0.9922 (p-values $< 10^{-16}$) for group B and rho = 0.9914 (p-value $< 10^{-16}$) for group A. (Fig. 2).

In addition, a second confirmation test was performed to assess the potential to estimate accurate β-values in the most significant differentially methylated CpGs when using a DNA pool strategy in EWAS. A comparison between all CpGs between group A and group B was performed and the most significant CpG sites ($n = 101$), p-val $< 10^{-05}$, were selected. Highly significant correlations (p-val $< 10^{-16}$) were also observed when analysing the 101 selected CpGs between group A and their pooled samples and between group B and their pooled samples (Group A: rho = 0.9808, p-value $< 10^{-16}$; Group B: rho = 0.9872, p-value $< 10^{-16}$) (Fig. 3).
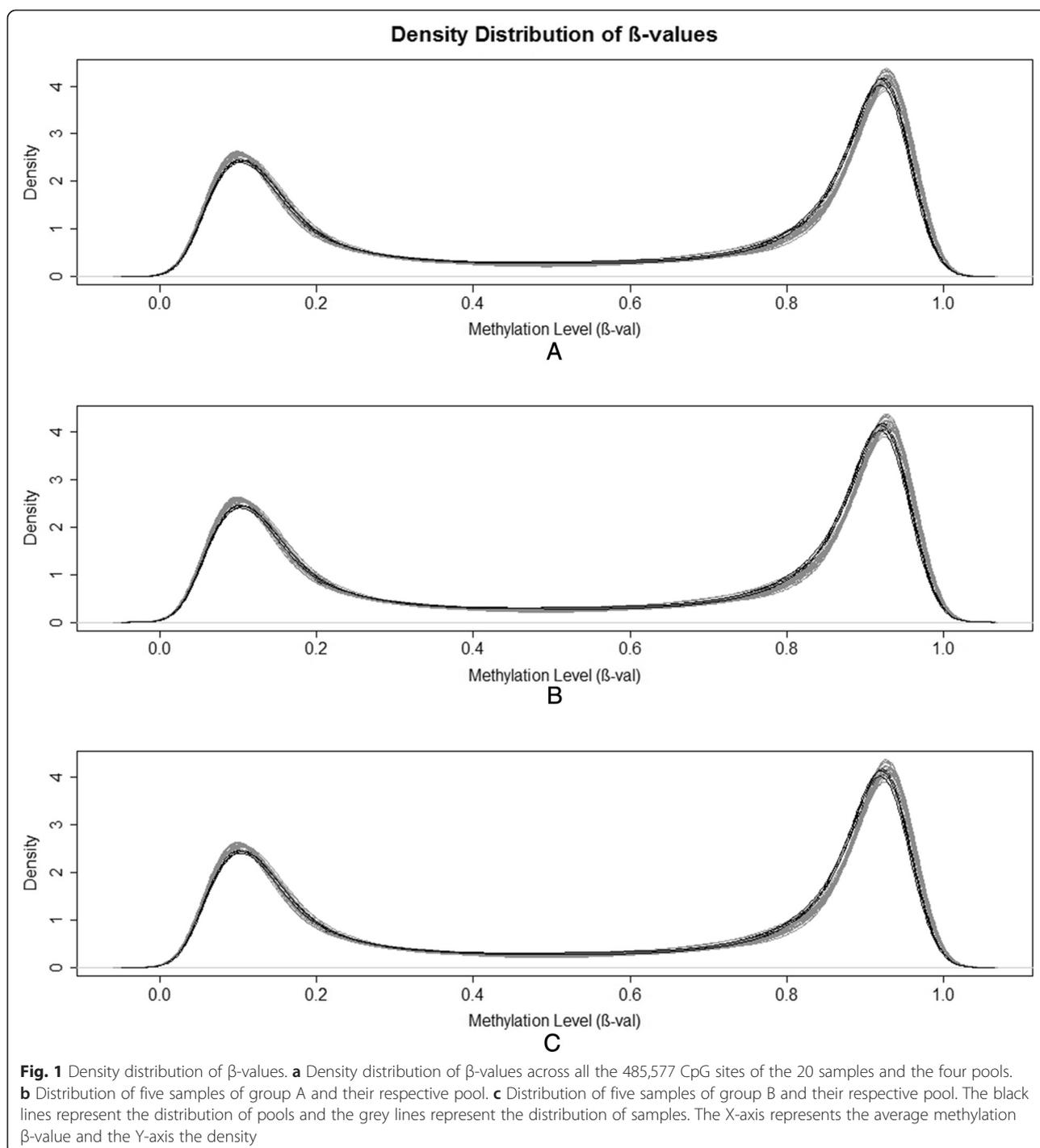
Sample size

Using values from the most significant DMC of pooled samples in the EWAS study, the optimum sample size to reach a 95 % statistical power and a 10^{-6} significance level, should be from 43 to 100 pooled samples per condition, considering Cohen's d effect sizes of 1.5 to 0.95 respectively.

The accuracy and reproducibility of DNA pools for methylation array, using the Illumina Infinium HumanMethylation450 BeadChip array, was investigated by comparing data obtained from individual samples and the same samples after they had been pooled.

Our data indicate that the DNA methylation profile (β-values of CpG sites) from the pooled DNA samples using array technology are highly consistent with those obtained from the individual samples, even when evaluating the most significant DMCs separately (Group A: rho = 0.9808, p-value $< 10^{-16}$; Group B: rho = 0.9872, p-value $< 10^{-16}$).
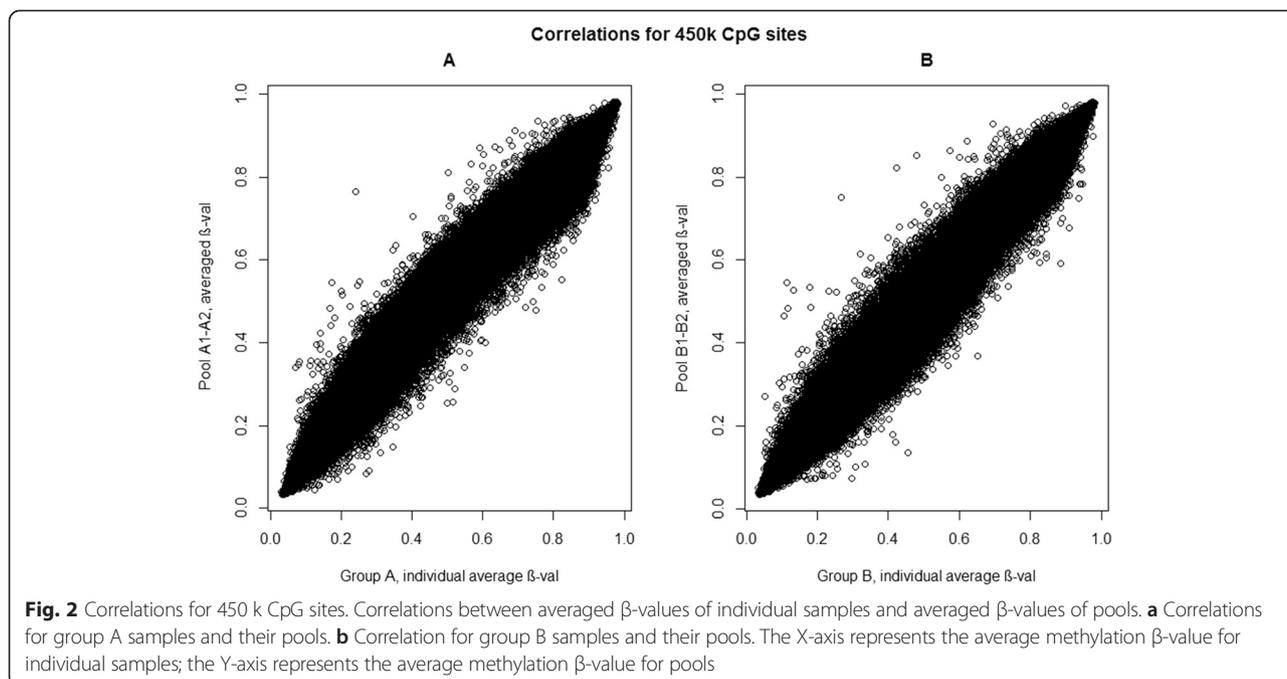
A previous study analysing pooling strategies in methylation studies demonstrated that pools could be an alternative technique when small amounts of DNA are available or when a reduction in cost is necessary to undertake the experiments. In the study, Docherty et al. showed a correlation between 89 individual samples and 4 pool samples in 205 CpG sites spanning 9 genomic regions using Sequenom EpiTYPER [19]. The overall correlation value in the study was 0.95 with a p-value $< 2.2 10^{-16}$, similar to the results that we observed. However, in our study we found that pooling strategies can be also performed assessing whole genomes in array-based EWAS experiments, analysing more than 450,000 CpG sites. This finding expands the possibilities of Genome Wide studies in epigenetics. In a pooling-based GWAS study, Pearson et al. demonstrated successful identification of published genetic susceptibility loci for some human

Gallego-Fabrega *et al. Clinical Epigenetics* (2015) 7:78

Page 3 of 8

Fig. 1 Density distribution of β-values. **a** Density distribution of β-values across all the 485,577 CpG sites of the 20 samples and the four pools. **b** Distribution of five samples of group A and their respective pool. **c** Distribution of five samples of group B and their respective pool. The black lines represent the distribution of pools and the grey lines represent the distribution of samples. The X-axis represents the average methylation β-value and the Y-axis the density

diseases: *APOE-ε4* in Alzheimer disease, *MAPT* in progressive supranuclear palsy and *TSPYL* in sudden infant death with dysgenesis of the testes syndrome (SIDDT) [15]. In EWAS we have yet to confirm whether previously reported genes can be found using pooling strategies. However, the higher correlation of the methylation levels between pools and individual samples indicates that the pooling strategies
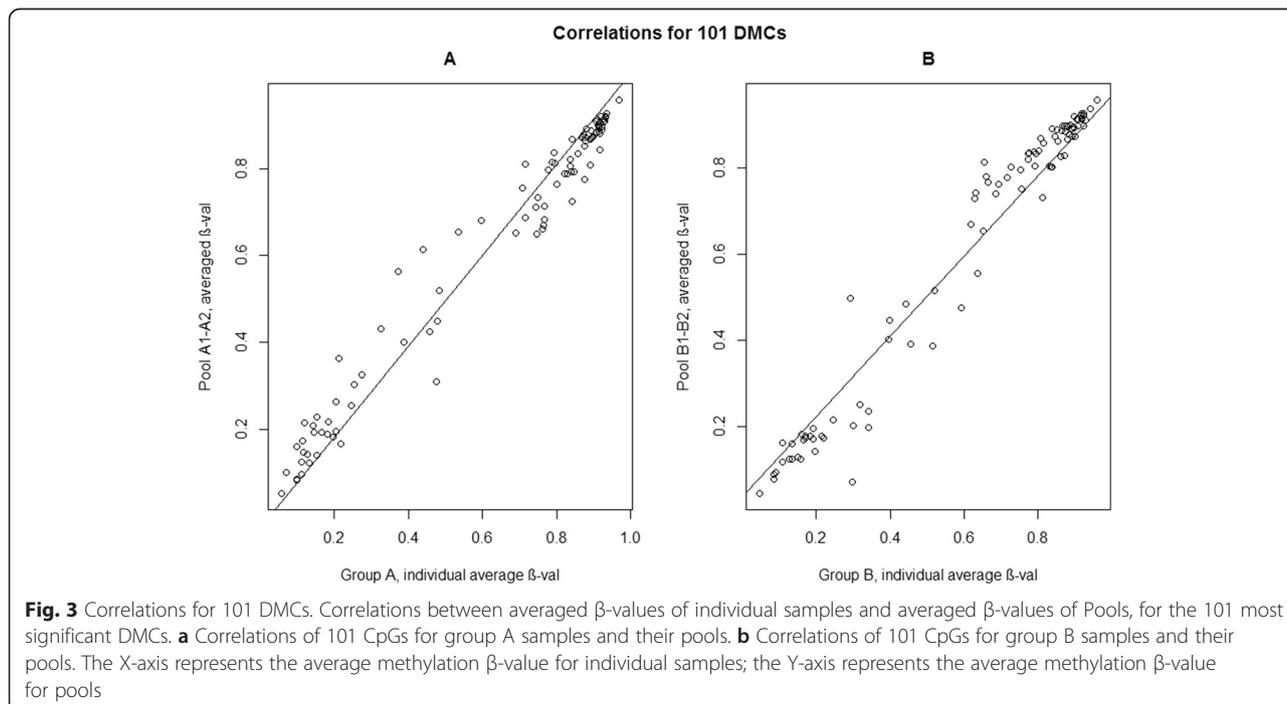
in EWAS are an accurate and interesting strategy to reduce time costs and DNA amount in such experiments.

Even though a DNA pooling strategy has important advantages, there are several drawbacks that have to be considered in the study design. Pool construction has to be really precise. DNA quantities have to be really accurate to assure that each sample in the pool provides equal

Gallego-Fabrega *et al. Clinical Epigenetics* (2015) 7:78

Page 4 of 8

Fig. 2 Correlations for 450 k CpG sites. Correlations between averaged β-values of individual samples and averaged β-values of pools. **a** Correlations for group A samples and their pools. **b** Correlation for group B samples and their pools. The X-axis represents the average methylation β-value for individual samples; the Y-axis represents the average methylation β-value for pools

quantities of DNA in order to minimize technical errors that may alter the estimated methylation levels [14, 20]. Only mean methylation levels, and not individual methylation data, can be obtained from pooled samples. In addition, adjusting for covariates is almost impossible, unless pooled samples are very homogeneous. Population stratification needs to be excluded. Furthermore,

the error rate tends to be higher in pooled samples compared to individual ones [21]. It is also important in the study design for EWAS with pooled samples to take into account the sample size needed to compute DMCs with confidence. According to the results obtained in our study, we suggest analysis of at least $n = 43$ pooled samples per condition in order to achieve a 10^{-06}

Fig. 3 Correlations for 101 DMCs. Correlations between averaged β-values of individual samples and averaged β-values of Pools, for the 101 most significant DMCs. **a** Correlations of 101 CpGs for group A samples and their pools. **b** Correlations of 101 CpGs for group B samples and their pools. The X-axis represents the average methylation β-value for individual samples; the Y-axis represents the average methylation β-value for pools

Gallego-Fabrega *et al. Clinical Epigenetics* (2015) 7:78

Page 5 of 8

significance level and 95 % statistical power, considering a Cohen's d effect size =1.5. However, this number may vary depending on a study's characteristics.

In summary, this is the first study that analyses a pooling strategy in EWAS approaches, it found that this strategy is an acceptable alternative to regular individual EWAS analysis, mainly in specific situations such as when lower quantities of DNA are available, or in studies with a limited budget.

Conclusions

The analysis of the data generated by 450,000 CpG sites across the whole genome in 20 individual samples demonstrates that DNA pooling strategies can be used to provide estimations of averaged DNA methylation state using the Illumina Infinium HumanMethylation450 BeadChip array. This approach may be useful to highlight genome regions to be studied in further epigenetic analysis, reducing the costs and the amount of DNA required.

Methods

Sample selection and pool construction

A total of 20 subjects from our biobank were selected. Of these 20 subjects, 10 were ischemic stroke patients with vascular recurrence (this selection was performed randomly from the patients with vascular recurrence) (group A). These patients were then matched one-to-one with 10 ischemic stroke patients without vascular recurrence (group B). The matching categories were age (±7 years), sex, TOAST classification [19] and recruitment hospital. Next, two pooled samples were constructed with samples from group A (PoolA1 and PoolA2), and two pooled samples were constructed with samples from group B (PoolB1 and PoolB2), as described in Fig. 4. PoolB1 included the matched samples of PoolA1 and correspondingly PoolB2 included the matched samples of PoolA2. All individuals were

Table 1 Population characteristics

		Total	Group A	Group B
Gender	*N*	20	10(50 %)	10 (50 %)
	Age	72,25 ± 8.4	72,5 ± 8.4	72 ± 8.7
	Male	16 (80 %)	8 (40 %)	8 (40 %)
	Female	4 (20 %)	2 (10 %)	2 (10 %)
TOAST[a]	Atherothrombotic	8 (40 %)	4 (20 %)	4 (20 %)
	Undetermined	4 (20 %)	2 (10 %)	2 (10 %)
	Unknown	2 (10 %)	1 (5 %)	1 (5 %)
	Lacunar	6 (30 %)	3 (15 %)	3 (15 %)

Demographic and clinical variables of the studied population
[a]TOAST classification of ischemic stroke

Caucasian, while 16 were males and 4 were females, mean age was 71 ± 8 years (Table 1).

DNA purification and sample pooling

Total genomic DNA was extracted from whole blood samples using the Gentra Puregene Blood Kit (Quiagen, Hilden, Germany) following the manufacturer's instructions. The samples were maintained at −20 °C until the EWAS analysis.

DNA concentrations for each subject were determined individually, by measuring ultraviolet (UV) light absorption at 260 nm, with NanoDrop 2000 UV–vis Spectrophotometer (Thermo Scientific, Redwood City, CA, USA). Adapting the instructions of previous DNA pooling protocols [14, 15], each sample was diluted to 40 ng/ul, and their DNA concentration was measured again to verify that all samples provide the same amount of DNA to the pools. Samples with DNA concentration variations higher than 40 ng/μl ±4 were discarded and diluted again. Individual DNA samples were then added to their respective pool (4 μl at 40 ng/μl of each sample). Once each pool was generated, the DNA concentrations were requantified twice with NanoDrop to assure that the final concentration of the pool was as expected (40 ng/μl). If any discrepancy was found (> ± 4 ng/μl), the pool was
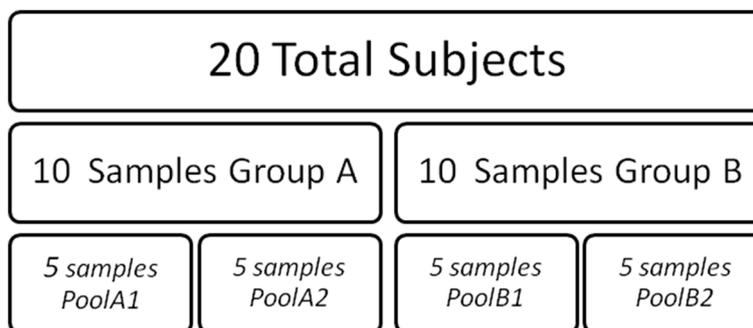

Fig. 4 Study design. A total of 20 subjects from the biobank study were selected for this study, 10 of them form group A and 10 form group B. Two pools were created from group A, with 5 samples in each one (PoolA1 and PoolA2), and two more pools of 5 samples were created from group B (PoolB1 and PoolB2)

Gallego-Fabrega *et al. Clinical Epigenetics* (2015) 7:78

Page 6 of 8

generated again repeating all steps from sample DNA measures. Only when the final pool concentration was 40 ng/µl ±4 ng/µl and the total volume was 20 µl as expected was the EWAS analysis started. A graphical description of the procedure can be found in Fig. 5.

Epigenome wide association analysis

Genome-wide DNA methylation was assessed using the Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, Ca). This chip-based study quantitatively measures more than 450,000 CpG sites at single nucleotide resolution with a 99 % coverage of RefSeq Genes.

A Quality Control (QC) of all samples was performed as a first step to check DNA integrity using Invitrogen E-Gel 1 % Agarose Gels. The DNA samples showed no fragmentation or poor quality.

Genomic DNA from the 20 samples and the 4 pools was bisulphite converted using the Zymo EZ DNA Methylation[TM] Kit (Zymo Research, Orange, Ca) following the manufacturer's instructions, but with alternative incubation conditions suggested for the Illumina Infinium Methylation Assay. All samples were processed in a single working batch using the Illumina Infinium MSA4 protocol, which includes amplification, fragmentation, hybridization and BeadChip scanning.

For QC, the fluorescence data generated for each CpG locus was analysed with the Illumina GenomeStudio software package. Samples and CpG sites with fluorescence detection p-values > 0.05 were removed [22]. This p-vaule is the detection p-value that represents the confidence that a given methylation level on a CpG site can be considered to have been detected.

Quality control and normalization

All pre-processing, correction and normalization steps were implemented using the R computing environment (versions 2.15.1 and 3.0.1) with Bioconductor packages. Plots were produced using R functions. The pipeline was
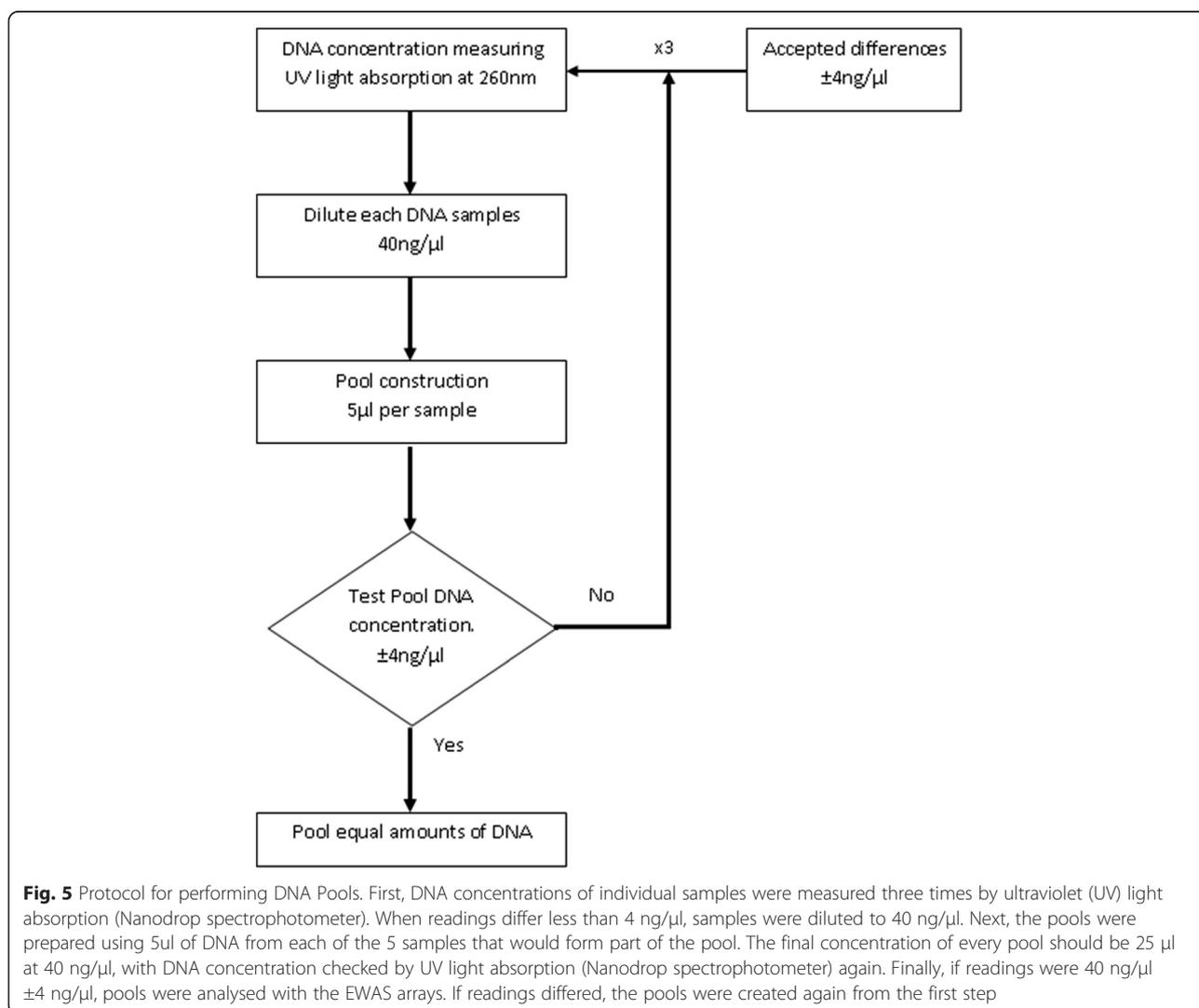

Fig. 5 Protocol for performing DNA Pools. First, DNA concentrations of individual samples were measured three times by ultraviolet (UV) light absorption (Nanodrop spectrophotometer). When readings differ less than 4 ng/µl, samples were diluted to 40 ng/µl. Next, the pools were prepared using 5ul of DNA from each of the 5 samples that would form part of the pool. The final concentration of every pool should be 25 µl at 40 ng/µl, with DNA concentration checked by UV light absorption (Nanodrop spectrophotometer) again. Finally, if readings were 40 ng/µl ±4 ng/µl, pools were analysed with the EWAS arrays. If readings differed, the pools were created again from the first step

Gallego-Fabrega *et al. Clinical Epigenetics* (2015) 7:78

Page 7 of 8

Table 2 R packages and instructions. Specific instructions used from each R package

R package	Instruction	Description
methylumi	methylumiR	Load Illumina methylation data into a MethyLumiSet object.
minfi	densityPlot	Density plots of methylation Beta values.
watRmelon	pfilter	Filter data sets based on bead count and detection p-values
minfi	mdsPlot	Multi-dimensional scaling (MDS) plots showing a 2-d projection of distances between samples.
watRmelon	dasen	Calculate normalized betas from Illumina 450 K methylation arrays.
lumi	estimateBetas	Estimate methylation Beta-value matrix from eSet-class object (include methylated and unmethylated probe intensities)

a sequence of R scripts adapted from the methylumi [23] (version 2.6.1), lumi [24] (version 2.12.0), watermelon [25] (version 1.0.3) and minfi [22] (version 1.6.0) packages. The instructions that were used are shown in Table 2.

Prior to the identification of differentially methylated CpG sites, data was pre-processed using a non-specific filter step. This step consists of removing CpG sites with detection p-value ≥ 0.05 in more than 1 % of the samples. Samples with detection p-value ≥ 0.05 in more than 1 % of the CpG sites, and CpG sites with beadCount < 3 in 5 % of samples [16]. CpG sites containing documented single nucleotide polymorphisms (SNPs) were also removed [26]. Multidimensional scaling (MDS) plots were used to evaluate gender outliers based on chromosome X data, where males and females were separated into two distinct clusters. An MDS plot was also used to check for unknown population structures, inside the sample. Then, CpG sites on the X and Y chromosomes were removed [8]. Finally, a subset quantile normalization was performed using a background adjustment between-array normalization and a dye bias correction, following previous recommendations [27, 28].

Statistical analysis

All statistical analysis was also performed using R (version 3.0.1). The accuracy of DNA methylation level estimations from pooled DNA was assessed with a Spearman's correlation, for non-parametric samples, between the β-values of each pool and the averaged β-values of the individual samples included in each pool [19].

We also performed a Spearman's correlation between the β-values of the 101 most differentially methylated CpGs (DMCs) found in individual samples (Group A vs. Group B) and the β-values of the same CpG sites in pools. Differentially methylated CpG sites were determined by the Mann–Whitney U-test for non-parametric

samples using the β-values, p-val $< 10^{-06}$ adapted from Rakyan VK el al. [4]. The DMCs analysis was performed comparing group A samples ($n = 10$) against group B samples ($n = 10$).

Minimum sample size needed for pool analysis in EWAS was calculated using the *pwr* package [29] with implemented power analysis as outlined by J. Cohen, 1988.

Ethical considerations

Ethical approval has been obtained from the ethical committee of the Vall d'Hebron Hospital (PR(AG) 03/2007). All patients were provided with oral and written information about the project, and each participant signed the informed consent for the study.

Author details
[1]Stroke pharmacogenomics and genetics, Fundació Docència i Recerca Mutua Terrassa, Hospital Universitari Mútua de Terrassa, C/ Sant Antoni 19, 08221 Terrassa, Barcelona, Spain. [2]School of Medicine, University of Barcelona, Barcelona, Spain. [3]Neurovascular Research Laboratory, Institut de Recerca, Universitat Autònoma de Barcelona, Hospital Vall d'Hebron, Barcelona, Spain. [4]Servicio de Neurología, Hospital Universitari Mútua Terrassa, Terrasa, Barcelona, Spain. [5]School of Healthcare Science, Manchester Metropolitan University, Manchester, UK.

References
1. Langevin SM, Kelsey KT. The fate is not always written in the genes: epigenomics in epidemiologic studies. Environ Mol Mutagen. 2013;54:533–41.
2. Gopalakrishnan S, Van Emburgh BO, Robertson KD. DNA methylation in development and human disease. Mutat Res. 2008;647:30–8.
3. Jiang Y-H, Bressler J, Beaudet AL. Epigenetics and human disease. Annu Rev Genomics Hum Genet. 2004;5:479–510.
4. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12:529–41.
5. Bock C. Epigenetic biomarker development. Epigenomics. 2009;1:99–110.
6. Feinberg AP. Genome-scale approaches to the epigenetics of common human disease. Virchows Arch. 2010;456:13–21.
7. Verma M. Epigenome-Wide Association Studies (EWAS) in Cancer. Curr Genomics. 2012;13:308–13.

Gallego-Fabrega *et al. Clinical Epigenetics* (2015) 7:78

Page 8 of 8

8. Shen J, Wang S, Zhang Y-J, Wu H-C, Kibriya MG, Jasmine F, et al. Exploring genome-wide DNA methylation profiles altered in hepatocellular carcinoma using Infinium HumanMethylation 450 BeadChips. Epigenetics. 2013;8:34–43.

9. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, et al. Aging effects on DNA methylation modules in human brain and blood tissue. Genome Biol. 2012;13:R97.

10. Johnson AA, Akman K, Calimport SRG, Wuttke D, Stolzing A, de Magalhães JP. The role of DNA methylation in aging, rejuvenation, and age-related disease. Rejuvenation Res. 2012;15:483–94.

11. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, et al. DNA methylation as a long-term biomarker of exposure to tobacco smoke. Epidemiology. 2013;24:712–6.

12. Flom JD, Ferris JS, Liao Y, Tehranifar P, Richards CB, Cho YH, et al. Prenatal smoke exposure and genomic DNA methylation in a multiethnic birth cohort. Cancer Epidemiol Biomarkers Prev. 2011;20:2518–23.

13. Gupta R, Nagarajan A, Wajapeyee N. Advances in genome-wide DNA methylation analysis. Biotechniques. 2010;49:iii–xi.

14. Sham P, Bader JS, Craig I, O'Donovan M, Owen M. DNA Pooling: a tool for large-scale association studies. Nat Rev Genet. 2002;3:862–71.

15. Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, Homer N, et al. Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. Am J Hum Genet. 2007;80:126–39.

16. Gaj P, Maryan N, Hennig EE, Ledwon JK, Paziewska A, Majewska A, et al. Pooled sample-based GWAS: a cost-effective alternative for identifying colorectal and prostate cancer risk variants in the Polish population. PLoS One. 2012;7:e35307.

17. Xu X, Su S, Barnes VA, De Miguel C, Pollock J, Ownby D, et al. A genome-wide methylation study on obesity: differential variability and differential methylation. Epigenetics. 2013;8:522–33.

18. Häsler R, Feng Z, Bäckdahl L, Spehlmann ME, Franke A, Teschendorff A, et al. A functional methylome map of ulcerative colitis. Genome Res. 2012;22:2130–7.

19. Docherty SJ, Davis OSP, Haworth CMA, Plomin R, Mill J. Bisulfite-based epityping on pooled genomic DNA provides an accurate estimate of average group DNA methylation. Epigenetics Chromatin. 2009;2:3.

20. Norton N, Williams NM, O'Donovan MC, Owen MJ. DNA pooling as a tool for large-scale association studies in complex traits. Ann Med. 2004;36:146–52.

21. Teumer A, Ernst FD, Wiechert A, Uhr K, Nauck M, Petersmann A, et al. Comparison of genotyping using pooled DNA samples (allelotyping) and individual genotyping using the affymetrix genome-wide human SNP array 6.0. BMC Genomics. 2013;14:506.

22. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9.

23. Davis S, Du P, Bilke S, Triche T J and BM. methylumi: Handle Illumina methylation data. R Packag version 2100. 2014.

24. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008;24:1547–8.

25. Schalkwyk LC, Pidsley R, Wong CC, Touleimat wfcbN, Defrance M TA and MJ. wateRmelon: Illumina 450 methylation array normalization and metrics. R Packag version 140. 2013.

26. Price ME, Cotton AM, Lam LL, Farré P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics Chromatin. 2013;6:4.

27. Touleimat N, Tost J. Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics. 2012;4:325–41.

28. Pidsley R, Wong CC Y, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics. 2013;14:293.

29. Basic Functions for Power Analysis. [http://cran.r-project.org/web/packages/pwr/pwr.pdf].