Epigenetics 8:10, 1123–1124; October 2013; © 2013 Landes Bioscience

Report on the 2nd Annual Infinium Humanmethylation450 Array Workshop

15 April 2013 QMUL, London, UK

Robert Lowe[1] and Tiffany Morris[2]

[1]Barts and the London School of Medicine and Dentistry; Queen Mary University of London; London, UK; [2]University College London; Cancer Institute; London, UK

Keywords: 450k, methylation, epigenetics, workshop, report, London, Illumina, HumanMethylation450, April

The Illumina Infinium HumanMethylation450 BeadChip—the successor to their hugely popular HumanMethylation27 BeadChip—is arguably the most prevalent platform for large-scale studies of DNA methylome analysis. After the success of last year's meeting[1] that discussed initial analysis strategies for this then-new platform, this year's meeting (held at Queen Mary, University of London) included the presentation of now established pipelines and normalization methods for data analysis, as well as some exciting tools for down-stream analysis. The importance of defining cell composition was a new topic mentioned by most speakers. The epigenome varies between cell types and insuring that methylation differences are related to sample treatment and not a differing cell population is essential. The meeting was attended by 215 computational and bench scientists from 18 countries. There were 11 speakers, a small poster session, and a discussion session. Talks were recorded and are now freely available at http://www.illumina.com/applications/epigenetics/array-based_methylation_analysis/methylation-array-analysis-education.ilmn

Approaches to 450k Analysis

Christoph Bock (CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Austria) was one of two invited speakers and gave an overview of DNA methylation data in general terms as well as the number of different assays available. Dr Bock highlighted the 450k array's accuracy and similarity to genotyping applications in addition to having a quick and easy analysis workflow compared with sequencing based methods and to this end briefly presented a 450k analysis tool developed in his group, RnBeads (http://rnbeads.bioinf.mpi-inf.mpg.de). Other tools mentioned, in which Dr Bock is a contributor, included EpiExplorer, EpiGRAPH, BiQ Analyzer, and MethMarker. Finally, Dr Bock presented GoDMC (Genetics of DNA Methylation Consortium) that facilitates genome-wide association studies of DNA methylation. This consortium has been organized to bring together researchers studying the genetic basis of DNA methylation and provide a centralized hub for coordinating analysis, summary statistics, replication, and meta-analysis.

In addition to RnBeads, a number of other software packages have been developed for 450k data analysis. Chloe Wong (Kings College London, UK) presented their recently published package, Watermelon,[2] available from Bioconductor. Watermelon provides a number of QC steps including the use of SNP probes for sample relatedness and identification of possible sample mix-ups. The package provides a wide variety of normalization options and they have compared 15 types of preprocessing methods in which they found their own Dasen method as the most effective. However, they encourage users to experiment and determine the method best suited to their data.

Tiffany Morris (Cancer Institute, University College London, UK) presented the recently released ChAMP package, an R-package available for download at http://www2.cancer.ucl.ac.uk/medicalgenomics/champ/. This package offers novel methods for downstream analysis. These include a function for batch effect estimation using singular-value decomposition (SVD) that highlights technical variation; a new way to call differentially methylated regions (DMRs) using a feature-oriented dynamic window that aims to capture neighboring significant probes; and also a function for estimating copy number aberrations (CNAs) in 450k data.

Kasper Hansen (John Hopkins Bloomberg School of Public Health, USA) presented the minfi package available from Bioconductor. Dr Hansen gave an overview of the currently available functions for QC, normalization, and also mentioned functions that would be released in the near future for predicting the sex of samples to highlight mislabeled samples, a function to estimate cell composition and a DMR caller. Dr Hansen particularly emphasized the importance of considering sample mix-ups, especially in large EWAS studies as up to 10% of samples can be mislabeled.

Data Integration and Methylation Studies

Kim Siegmund (University of Southern California, USA) was the 2nd invited speaker and she discussed her implementation of the methylumi package available from Bioconductor. Dr Siegmund focused on the importance of background correction

Correspondence to: Robert Lowe and Tiffany Morris;
Email: r.lowe@qmul.ac.uk and tiffany.morris@ucl.ac.uk
Submitted: 08/01/13; Accepted: 08/07/13
http://dx.doi.org/10.4161/epi.26073

(of both Type I and Type II probes) and dye bias adjustment (of Type II probes) and emphasized the impact of background noise on β values, principally for studies where small changes are expected. In these studies it is essential to properly process signals. Dr Siegmund has found that the removal of background noise and adjustment for dye bias removes the commonly observed trend of increasing signals across an array. As such, Dr Siegmund is confident that there is now a strong understanding of normalization methods, however, it is important to next focus on differentiating cell types. She ended by giving an overview of publically available data sets from The Cancer Genome Atlas (TCGA).

Francesco Marabita (Karolinska Institute, Sweden) presented recently published work[3] of his evaluation of six different analysis pipelines that differ in normalization method. A number of pipelines now exist and it is important to evaluate them. He used two unpublished and two published data sets to make the comparison and found quantile normalization and Beta Mixture Quantile dilation (BMIQ) to be the most effective methods.

Robert Lowe (Queen Mary University of London, UK) presented his recent work on a new public database for Illumina 450k data known as Marmal-Aid (http://marmal-aid.org/). This allows easy access to all publically available 450k data (currently 8000+ samples). The data are extracted from public repositories and a number of pre-processing steps are applied such as data imputation and normalization. He showed that samples cluster by tissue rather than batch that suggests methylation differences of >10% were detectable above the influence of batch.

Roderick Slieker (Leiden University Medical Center, The Netherlands) presented a method for characterizing tissue specific Differentially Methylated Regions (tDMRs) in two independent data sets of four peripheral tissues (blood, saliva, buccal swab, and hair follicles) and six internal tissues (liver, muscle, pancreas, subcutaneous fat, omentum, and spleen with paired blood). Of the tDMRs, 13% mapped to gene body CpG islands and 25% to CpG islands shores. Implementation of annotations that recently became available through ENCODE showed enrichment of tDMRs in DNase hypersensitive sites and transcription factor binding sites.

Amy Webster (Manchester University, UK) presented results in relation to her study on differential methylation related to response of etanercept in patients with rheumatoid arthritis. A huge advance in the treatment of rheumatoid arthritis (RA) has been the introduction of biologic drug therapies; however, 40% of patients fail to respond. They chose 24 patients to identify a methylation signature indicative of response and found nine probes that showed significant differences. While the sample size was small this preliminary data indicate a possible methylation biomarker of response.

The program was brought to a satisfying close following two talks given by representatives from Illumina: Bret Barnes (Illumina, USA) explained clearly the rationale behind the design of the 450k array and the requirement for a mixture of probe types, while Fraz Syed (Illumina, USA) discussed a whole genome bisulfite sequencing library preparation kit.

Discussion and Future Considerations

It is clear that analysis pipelines for the 450k have evolved over the past year. While last year's talks focused heavily on Type 1 and Type 2 differences, these appear less of a concern now. Details of the preprocessing methods are very important, particularly in EWAS studies and a number of effective methods are now available. One issue that was mentioned in a number of talks was the importance of knowing the cell type composition in an experiment. Dr Hansen highlighted this particularly effectively during his talk, illustrating the point by showing aging data where newborns have a significantly different blood composition to that of the elderly and hence correcting for this is essential before looking for changes in methylation between the two groups.

Readers wishing to keep informed of issues related to the Illumina 450k platform are encouraged to visit the actively-used web forum set up last year, http://groups.google.com/group/epigenomicsforum

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

This meeting was fully funded by Illumina (www.illumina.com) and Partek (www.partek.com) and supported by two EU-FP7 consortia: BLUEPRINT and IDEAL

References

1. Morris T, Lowe R. Report on the Infinium 450k methylation array analysis workshop: April 20, 2012 UCL, London, UK. Epigenetics 2012; 7:961-2; PMID:22722984; http://dx.doi.org/10.4161/epi.20941
2. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 2013; 14:293; PMID:23631413; http://dx.doi.org/10.1186/1471-2164-14-293
3. Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerström-Billai F, Jagodic M, Sundberg CJ, Ekström TJ, Teschendorff AE, Tegnér J, et al. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. Epigenetics 2013; 8:333-46; PMID:23422812; http://dx.doi.org/10.4161/epi.24008