Bioinformatics, 37(15), 2021, 2198–2200
doi: 10.1093/bioinformatics/btaa1060
Advance Access Publication Date: 26 December 2020
Applications Note

Genome analysis

MethPanel: a parallel pipeline and interactive analysis tool for multiplex bisulphite PCR sequencing to assess DNA methylation biomarker panels for disease detection

Phuc-Loi Luu[1,2,*,†], Phuc-Thinh Ong[3,†], Tran Thai Huu Loc[4], Dilys Lam[1], Ruth Pidsley[1,2], Clare Stirzaker[1,2,†] and Susan J. Clark [1,2,*,†]

[1]Epigenetics Research Laboratory, Genomics and Epigenetics Theme, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia, [2]St Vincent's Clinical School, UNSW, Sydney, NSW 2010, Australia, [3]Center for Population Health Sciences, Hanoi University of Public Health, Hanoi 100000, Vietnam and [4]School of Medicine, Vietnam National University, Ho Chi Minh City 70000, Vietnam

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors and last two authors should be regarded as Joint Authors.
Associate Editor: Peter Robinson

Received on August 26, 2020; revised on November 1, 2020; editorial decision on November 16, 2020; accepted on December 13, 2020

Abstract

Summary: DNA methylation patterns in a cell are associated with gene expression and the phenotype of a cell, including disease states. Bisulphite PCR sequencing is commonly used to assess the methylation profile of genomic regions between different cells. Here we have developed MethPanel, a computational pipeline with an interactive graphical interface to rapidly analyse multiplex bisulphite PCR sequencing data. MethPanel comprises a complete analysis workflow from genomic alignment to DNA methylation calling and supports an unlimited number of PCR amplicons and input samples. MethPanel offers important and unique features, such as calculation of an epipolymorphism score and bisulphite PCR bias correction capabilities, and is designed so that the methylation data from all samples can be processed in parallel. The outputs are automatically forwarded to a shinyApp for convenient display, visualization and remotely sharing data with collaborators and clinicians.
Availabilityand implementation: MethPanel is freely available at https://github.com/thinhong/MethPanel.
Contact: s.clark@garvan.org.au
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Methylation of CpG dinucleotides plays a crucial role in the regulation of mammalian gene expression and therefore in the development of different pathologies. As such, analysis of DNA methylation patterns allows researchers to explore biological aging, disease genesis and progression (Dor and Cedar, 2018) and in particular precision oncology (Weichenhan *et al.*, 2020). Since DNA methylation states are highly consistent and stable they are increasingly used in clinical diagnostics (Gai and Sun, 2019). A recent community-wide benchmarking study comparing DNA methylation assays showed that bisulphite PCR sequencing had the best all-round performance for measuring DNA methylation in clinical samples with low-input DNA (Bock *et al.*, 2016). Moreover, bisulphite PCR sequencing can be multiplexed to target multiple regions simultaneously, which is a critical requirement for clinical applications (Lu *et al.*, 2017). Multiplex bisulphite PCR sequencing can also accurately measure methylation in archival, formalin-fixed paraffin-embedded tissue (FFPET) DNA (Korbie *et al.*, 2015) and may therefore be applicable to circulating cell free DNA (cfDNA) in plasma (Gai and Sun, 2019).

Implementation of DNA methylation markers in clinical practice requires the development of user-friendly, large-scale sample analysis and a diversity of visualization options, which are not always met by current tools. For instance, *BSPAT* (Hu *et al.*, 2015) and *ampliMethProfiler* (Scala *et al.*, 2016) can process raw data, but lack an interactive interface. *Methpat* can produce interactive plots, but only focuses on visualization of DNA methylation patterns and requires considerable manual labour (Wong et al., 2016). *TABSAT* performs all levels of analysis and has an interactive interface, but is optimized for Ion Torrent and Illumina MiSeq platforms (Pabinger *et al.*, 2016). *EPIC-TABSAT* can generate multiple interactive plots and statistics, but is reported to work best with a maximum of 150 targets and 50 samples (Krainer *et al.*, 2019).

Following our recent comprehensive protocol for multiplex bisulphite PCR sequencing (Lam *et al.*, 2020), we now present

MethPanel, an easy-to-use tool for analysis and visualization that has been optimized for multiplex bisulphite PCR sequencing data. The computational tool covers a complete analysis workflow from alignment to DNA methylation calling. *MethPanel* does not limit the number of targets or samples, and importantly offers rarely provided features, such as calculation of an epipolymorphism score and correction for bisulphite PCR bias, that are especially relevant for tumour samples and cfDNA analyses. *MethPanel* can be run on personal computers or high performance computing machines, which allows multiple samples to be performed in parallel in a resumable manner. Outputs are automatically forwarded to a shinyApp for visualization where users can effortlessly customize and remotely share their full data and plots.

2 Results

2.1 Workflow and input data

The workflow of *MethPanel* was built with *Bpipe* (Sadedin *et al.*, 2012) in the Linux operating system (CentOS 8) as it allows all modules to run automatically, in parallel by samples to generate results in separate directories (see Upstream analysis in Fig. 1a and Supplementary Part S1). A key feature of this pipeline is that data analyses can be paused and resumed at any time. For the upstream analysis *MethPanel* requires the raw sequencing FASTQ files, as well as configuration input files: (i) an annotation file of the panel; (ii) a plain text file listing all sample names; and (iii) a config file of trimming and alignment parameters. *MethPanel* supports the human reference genome version hg18, hg19 and hg38.

2.2 Short sequence quality control and alignment

In *MethPanel* FASTQ files are trimmed with *TrimGalore* (https://github.com/FelixKrueger/TrimGalore) to produce high quality reads and to report the percentage of trimmed and remaining reads for each sample. *Bismark* (Krueger and Andrews, 2011) is used in the alignment module with user-defined parameters. To reduce computational memory requirement, *MethPanel* only maps the trimmed reads to the amplicon regions supplied in the user's annotation file. Internet connection is required to obtain the indexed reference of the genome. A report is generated that describes the number of initial reads supplied, number of mapped reads, number of filtered mapped reads (base quality >30 Phred score and length >20 bp, can be set in the configuration file), sequencing coverage of CpGs, the number of dropout CpGs (coverage <100×, can be set in the configuration file) and dropout amplicons (those without sufficient coverage of all CpGs). Average methylation levels of CpGs and non-CpG cytosines for each sample are also reported. To filter reads that failed bisulphite conversion, users can define the number of non-CpG cytosine residues permitted per read in the config file.

2.3 DNA methylation analysis, pattern calling and PCR bias correction

MethPanel performs DNA methylation calling for each sample and compiles the results into a table of methylation values for all samples. In this table, each row represents a CpG site, each column represents a sample and includes DNA methylation ratio, the number of methylated reads and coverage. *MethPanel* then performs calling of DNA methylation patterns (rather than just average methylation across an amplicon) and computes an epipolymorphism score for each sample. For k CpGs in an amplicon, all combinations of 2^k patterns are shown in a matrix plot. Epipolymorphism score is computed as $p_{epipoly} = 1 - \sum p_i^2$, of which p_i is the percentage of pattern i (Landan *et al.*, 2012). An optional step is to compute PCR-bias values based on the equation:

$$b = \frac{y(100 - x)}{x(100 - y)} \text{ (Warnecke } et\ al.,\ 1997)$$

In the equation, x corresponding to the expected methylation value and y to the proportion of methylated DNA observed in the 'methylated-control' DNA samples.

With the assumption that the uncorrected curve is hyperbolic curve, DNA methylation is corrected as:

$$y_{\text{corrected}} = \frac{(100y_0 - 100y)}{(by - by_1 + y_0 - y)}$$

(Moskalev *et al.*, 2011), where y_0 and $y_1$are the expected minimum and maximum DNA methylation, respectively.

2.4 Visualization with shinyApp

Table outputs in the previous stages are automatically forwarded to a shinyApp server accompanying the tool. Each user will be given an account with a password to ensure the privacy of their projects. The shinyApp Client displays data across seven tabs, summarized as follows (see shinyApp client in Fig. 1b and Supplementary Part S2): (i) Trimmed FASTQ metrics table. A detailed version of this table can be downloaded for further examination. (ii) Alignment metrics table, with accompanying visualizations. (iii) DNA methylation tab comprising three pages providing: (a) an overview of coverage and DNA methylation ratio with heatmap visualization, (b) a summary of DNA methylation at single amplicons and all amplicons with boxplot visualization and (c) a comparison of DNA methylation between groups. (iv) Heatmaps of within-sample DNA methylation patterns. (v) Plots of epipolymorphism scores for all amplicons. (vi) Plots for bias correction, if applicable. (vii) Bias-corrected data displayed and analysed in the same format as in the DNA methylation tab.

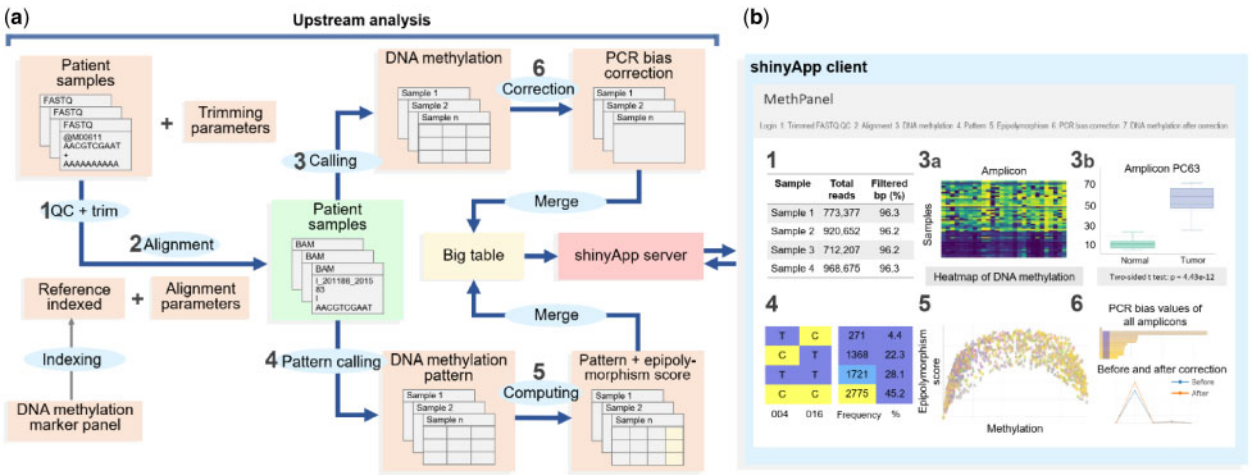

Fig. 1. *MethPanel* workflow: (**a**) upstream analysis; (**b**) shinyApp client summary

3 Conclusion

We have successfully developed *MethPanel*, a user-friendly visualization tool to analyse multiplex bisulphite PCR sequencing data with shinyApp interactive plots. The shinyApp server also centralizes the data making for storage management.

Acknowledgements

The authors thank Jenny Song and Wenjia Qu from the Clark Lab, who performed the multiplex PCR methylation sequencing assays that were used in this paper to demonstrate the utility of *MethPanel*.

Funding

This work was supported by National Health and Medical Research Council (NHMRC) Fellowship [1063559 to S.J.C.], Cancer Institute NSW Translational Program Grant [TPG172146 to S.J.C.], Cancer Council NSW [RG-18-09 to R.P. and S.J.C.], National Breast Cancer Foundation IIRS Grant [NBCF-IIRS-18-137] and National Foundation and Medical Research and Innovation (NFMRI) grant (C.S.) and computational resources provided by the Australian Government through NCI Raijin under the National Computational Merit Allocation Scheme 2020, project lv15 (S.J.C., P.-L.L.).

Conflict of Interest: none declared.

References

Bock,C. *et al.* (2016) Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nat. Biotechnol.*, **34**, 726–737.

Dor,Y. and Cedar,H. (2018) Principles of DNA methylation and their implications for biology and medicine. *Lancet (London, England)*, **392**, 777–786.

Gai,W. and Sun,K. (2019) Epigenetic biomarkers in cell-free DNA and applications in liquid biopsy. *Genes*, **10**, 32.

Hu,K. *et al.* (2015) BSPAT: a fast online tool for DNA methylation co-occurrence pattern analysis based on high-throughput bisulfite sequencing data. *BMC Bioinformatics*, **16**, 220.

Korbie,D. *et al.* (2015) Multiplex bisulfite PCR resequencing of clinical FFPE DNA. *Clin. Epigenet.*, **7**, 28.

Krainer,J. *et al.* (2019) EPIC-TABSAT: analysis tool for targeted bisulfite sequencing experiments and array-based methylation studies. *Nucleic Acids Res.*, **47**, W166–W170.

Krueger,F. and Andrews,S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571–1572.

Lam,D. *et al.* (2020) Comprehensive evaluation of targeted multiplex bisulphite PCR sequencing for validation of DNA methylation biomarker panels. *Clin. Epigenet.*, **12**, 90.

Landan,G. *et al.* (2012) Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat. Genet.*, **44**, 1207–1214.

Lu,J. *et al.* (2017) PrimerSuite: a high-throughput web-based primer design program for multiplex bisulfite PCR. *Sci. Rep.*, **7**, 41328.

Moskalev,E.A. *et al.* (2011) Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression. *Nucleic Acids Res.*, **39**, e77.

Pabinger,S. *et al.* (2016) Analysis and visualization tool for targeted amplicon bisulfite sequencing on ion torrent sequencers. *PLoS One*, **11**, e0160227.

Sadedin,S.P. *et al.* (2012) Bpipe: a tool for running and managing bioinformatics pipelines. *Bioinformatics*, **28**, 1525–1526.

Scala,G. *et al.* (2016) ampliMethProfiler: a pipeline for the analysis of CpG methylation profiles of targeted deep bisulfite sequenced amplicons. *BMC Bioinformatics*, **17**, 484.

Warnecke,P.M. *et al.* (1997) Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res.*, **25**, 4422–4426.

Weichenhan,D. *et al.* (2020) Epigenomic technologies for precision oncology. *Semin. Cancer Biol.*, in press.

Wong,N.C. *et al.* (2016) MethPat: a tool for the analysis and visualisation of complex methylation patterns obtained by massively parallel sequencing. *BMC Bioinformatics*, **17**, 98.