



Editorial: Computational Methods for Analysis of DNA Methylation Data

Pietro Di Lena^{1*}, Christine Nardini^{2*} and Matteo Pellegrini^{3*}

¹Department of Computer Science and Engineering, University of Bologna, Bologna, Italy, ²CNR IAC "Mauro Picone", Consiglio Nazionale delle Ricerche (CNR), Rome, Italy, ³Department of Molecular, Cell and Developmental Biology, UCLA, Los Angeles, CA, United States

Keywords: DNA methylation, DNA methylation age, epigenetic clocks, copy number variations, cell-type deconvolution, methylation pattern imputation

Editorial on the Research Topic

Editorial: Computational Methods for Analysis of DNA Methylation Data

DNA methylation is among the most studied epigenetic modifications in eukaryotes. The interest in DNA methylation stems from its role in development, as well as its well-established association with phenotypic changes. Particularly, there is strong evidence that methylation pattern alterations in mammals are linked to developmental disorders and cancer (Kulis and Esteller, 2010). Owing to its potential as a prognostic marker for preventive medicine, in recent years, the analysis of DNA methylation data has garnered interest in many different contexts of computational biology (Bock, 2012). As it typically happens with omic data, processing, analyzing and interpreting large-scale DNA methylation datasets requires computational methods and software tools that address multiple challenges. In the present Research Topic, we collected papers that tackle different aspects of computational approaches for the analysis of DNA methylation data. These manuscripts address novel computational solutions for copy number variation detection, cell-type deconvolution and methylation pattern imputation, while others discuss interpretations of well-established computational techniques.

Over the last 10 years, DNA methylation profiles have been successfully exploited to develop biomarkers of age, also referred to as epigenetic clocks (Bell et al., 2019). Epigenetic clocks accurately estimate both chronological and biological age from methylation levels. DNA methylation age and, most importantly, its deviation from chronological age have been shown to be associated with a variety of health issues. More recently, a second generation of epigenetic clocks has emerged. The new generation of clocks incorporates not only methylation profiles but also environmental variants, such as smoking and alcohol consumption, and they outperform the first generation in mortality prediction and prognosis of certain diseases. In our collection, the review by Chen et al. compares the first and second generation of epigenetic clocks that predict cancer risk and discusses pathways known to exhibit altered methylation in aging tissues and cancer.

Differentially methylated regions (DMRs), that is genomic regions that show significant differences in methylation levels across distinct biological and/or medical conditions (e.g., normal vs. disease), have been reported to be implicated in a variety of disorders (Rakyan et al., 2011). As a result, identifying DMRs is one of the most critical and fundamental challenges in deciphering disease mechanisms at the molecular level. Although DNA methylation patterns remain stable during normal somatic cell growth, alterations in genomic methylation may be caused by genetic alterations, or vice versa. However, standard DMR analysis often ignores whether

OPEN ACCESS

Edited and reviewed by:

Mark Borodovsky,
Georgia Institute of Technology,
United States

*Correspondence:

Pietro Di Lena
pietro.dilena@unibo.it
Christine Nardini
christine.nardini@cnr.it
Matteo Pellegrini
matteop@mcdb.ucla.edu

Specialty section:

This article was submitted to
Genomic Analysis,
a section of the journal
Frontiers in Bioinformatics

Received: 22 April 2022

Accepted: 25 May 2022

Published: 17 June 2022

Citation:

Di Lena P, Nardini C and Pellegrini M
(2022) Editorial: Computational
Methods for Analysis of DNA
Methylation Data.
Front. Bioinform. 2:926066.
doi: 10.3389/fbinf.2022.926066

methylation alterations should be viewed as a cause or an effect. Rhamani et al. discuss the effect of model directionality, i.e. whether the condition of interest (phenotype) may be affected by methylation or whether it may affect methylation, in differential methylation analyses at the cell-type level. They show that correctly accounting for model directionality has a significant impact on the ability to identify cell type specific differential methylation.

Different cell types exhibit DMRs at many genomic regions and such rich information can be exploited to infer underlying cell type proportions using deconvolution techniques. DNA methylation-based cell mixture deconvolution approaches can be classified into two main categories: reference-based and reference-free. While the latter are more broadly applicable, as they do not rely on the availability of methylation profiles from each of the purified cell types that compose a tissue of interest, they are also less precise. Reference-based approaches use DMRs specific to cell types (reference library) to determine the underlying cellular composition within a DNA methylation sample. The quality of the reference library has a big impact on the accuracy of reference-based approaches. Bell-Glenn et al. present RESET, a framework for reference library selection for deconvolution algorithms exploiting a modified version of the Dispersion Separability Criteria score, for the inference of the best DMRs composing the library, contributing to *de facto* standards (Koestler et al., 2016). In short, RESET does not require researchers to identify *a priori* the size of the reference library (number of DMRs), nor to rely on costly associated purified cells' mDNA profiles.

Within a cellular population, the methylation patterns of different cell types and at specific genomic locations are indicative of cellular heterogeneity. Alterations of such heterogeneity are predictive of development as well as prognostic markers of diseases. Computational methods that exploit heterogeneity in methylation patterns are typically constrained by partially observed patterns due to the nature of shotgun sequencing, which frequently generates limited coverage for downstream analysis. One possible solution to overcome such limitations is offered by Chang et al. presenting BSImp, a probabilistic based imputation method that uses local information to impute partially

observed methylation patterns. They show that using this approach they are able to recover heterogeneity estimates at 15% more regions with moderate sequencing depths. This should therefore improve our ability to study how methylation heterogeneity is associated with disease.

Finally, recent studies have shown how the associations between Copy Number Variations (CNVs) and methylation alterations offer a richer and hence more informative picture of the samples under study, in particular for tumor data characterized by large scale genomic rearrangements (Sun et al., 2018). Consequently, recent technological and methodological developments have enabled the possibility to measure CNVs from DNA methylation data. The main advantage of DNA methylation based CNV approaches is that they offer the possibility to integrate both genomic (copy number) and epigenomic (methylation) information. Mariani et al. propose MethylMasteR, an R software package that integrates DNA methylation-based CNV calling routines, facilitating standardization, comparison and customization of CNV analyses. This package, built into the Docker architecture to seamlessly manage dependencies, includes four of the most commonly used routines for this integrated analysis, ChAMP (Morris et al., 2014), SeSAmE (Zhou et al., 2018), Epicopy (Cho et al., 2019), plus a custom version of cnAnalysis450k (Knoll et al., 2017), overall enabling analysis of comparative results.

All the topics in this issue, although limited to specific aspects of DNA methylation analysis, highlight the importance of research in this field, the associated computational challenges and illustrate the significant impact that this type of data will likely have on preventive medicine.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We would like to thank all the reviewers who have participated in the review process of the articles submitted to this special issue.

REFERENCES

- Bell, C. G., Lowe, R., Adams, P. D., Baccarelli, A. A., Beck, S., Bell, J. T., et al. (2019). DNA Methylation Aging Clocks: Challenges and Recommendations. *Genome Biol.* 20, 249. doi:10.1186/s13059-019-1824-y
- Bock, C. (2012). Analysing and Interpreting DNA Methylation Data. *Nat. Rev. Genet.* 13, 705–719. doi:10.1038/nrg3273
- Cho, S., Kim, H. S., Zeiger, M. A., Umbricht, C. B., and Cope, L. M. (2019). Measuring DNA Copy Number Variation Using High-Density Methylation Microarrays. *J. Comput. Biol.* 26, 295–304. doi:10.1089/cmb.2018.0143
- Knoll, M., Debus, J., and Abdollahi, A. (2017). cnAnalysis450k: an R Package for Comparative Analysis of 450k/EPIC Illumina Methylation Array Derived Copy Number Data. *Bioinformatics* 33, 2266–2272. doi:10.1093/bioinformatics/btx156
- Koestler, D. C., Jones, M. J., Usset, J., Christensen, B. C., Butler, R. A., Kobor, M. S., et al. (2016). Improving Cell Mixture Deconvolution by Identifying Optimal DNA Methylation Libraries (IDOL). *BMC Bioinforma.* 17, 120. doi:10.1186/s12859-016-0943-7
- Kulis, M., and Esteller, M. (2010). DNA Methylation and Cancer. *Adv. Genet.* 70, 27–56. doi:10.1016/B978-0-12-380866-0.60002-2
- Morris, T. J., Butcher, L. M., Feber, A., Teschendorff, A. E., Chakravarthy, A. R., Wojdacz, T. K., et al. (2014). ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* 30, 428–430. doi:10.1093/bioinformatics/btt684
- Rakyan, V. K., Down, T. A., Balding, D. J., and Beck, S. (2011). Epigenome-Wide Association Studies for Common Human Diseases. *Nat. Rev. Genet.* 12, 529–541. doi:10.1038/nrg3000

- Sun, W., Bunn, P., Jin, C., Little, P., Zhabotynsky, V., Perou, C. M., et al. (2018). The Association between Copy Number Aberration, DNA Methylation and Gene Expression in Tumor Samples. *Nucleic Acids Res.* 46, 3009–3018. doi:10.1093/nar/gky131
- Zhou, W., Triche, T. J., Laird, P. W., and Shen, H. (2018). SeSAMe: Reducing Artifactual Detection of DNA Methylation by Infinium BeadChips in Genomic Deletions. *Nucleic Acids Res.* 46, e123. doi:10.1093/nar/gky691

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Di Lena, Nardini and Pellegrini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.