

<https://helda.helsinki.fi>

CpGmotifs : a tool to discover DNA motifs associated to CpG methylation events

Scala, Giovanni

2021-05-26

Scala , G , Federico , A & Greco , D 2021 , ' CpGmotifs : a tool to discover DNA motifs associated to CpG methylation events ' , BMC Bioinformatics , vol. 22 , no. 1 , 278 . <https://doi.org/10.1186/s12859-021-04191-8>

<http://hdl.handle.net/10138/332508>

<https://doi.org/10.1186/s12859-021-04191-8>

cc_by

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

SOFTWARE

Open Access



CpGmotifs: a tool to discover DNA motifs associated to CpG methylation events

Giovanni Scala¹, Antonio Federico^{2,3} and Dario Greco^{2,3,4*} 

*Correspondence:

dario.greco@tuni.fi

² Faculty of Medicine and Health Technology, Tampere University, 33014 Tampere, Finland
Full list of author information is available at the end of the article

Abstract

Background: The investigation of molecular alterations associated with the conservation and variation of DNA methylation in eukaryotes is gaining interest in the biomedical research community. Among the different determinants of methylation stability, the DNA composition of the CpG surrounding regions has been shown to have a crucial role in the maintenance and establishment of methylation statuses. This aspect has been previously characterized in a quantitative manner by inspecting the nucleotidic composition in the region. Research in this field still lacks a qualitative perspective, linked to the identification of certain sequences (or DNA motifs) related to particular DNA methylation phenomena.

Results: Here we present a novel computational strategy based on short DNA motif discovery in order to characterize sequence patterns related to aberrant CpG methylation events. We provide our framework as a user-friendly, shiny-based application, CpGmotifs, to easily retrieve and characterize DNA patterns related to CpG methylation in the human genome. Our tool supports the functional interpretation of deregulated methylation events by predicting transcription factors binding sites (TFBS) encompassing the identified motifs.

Conclusions: CpGmotifs is an open source software. Its source code is available on GitHub <https://github.com/Greco-Lab/CpGmotifs> and a ready-to-use docker image is provided on DockerHub at <https://hub.docker.com/r/grecolab/cpgmotifs>.

Keywords: DNA methylation, DNA motifs, DNA methylation signature, Transcription factors, R-Shiny

Background

DNA methylation aberration has been associated with several pathological conditions and, more recently, recognized as an important therapeutic target [1]. One of the most interesting aspects related to DNA methylation regards the ability of a CpG site to be more or less susceptible to changes in its methylation status. In fact, the propensity of CpG sites to de novo methylation is notably heterogeneous over the whole genome [2]. The reasons for such a discrepancy have been widely investigated. For instance, it has been demonstrated that a quantitative relationship exists between the DNA sequence composition and the stability of the associated methylation patterns [3, 4]. While this



quantitative relationship has been extensively studied [5, 6], the existence of qualitative features between a particular DNA sequence and its methylation status has not been fully explored. The concept of DNA motif can be used to associate CpG methylation changes to their DNA context [1]. The usage of DNA motifs allows researchers to easily explore families of closely related DNA sequences that are over-represented in a set of sequences of interest. Such an approach has been used in [7] to identify DNA motifs that regulate DNA methylation, in [8] to identify transcription factor-based epigenetic signatures in neural cells and in [1] DNA sequences associated to changes in methylation patterns in different cancer types. The concept of methylation related motifs analysis has already been implemented in several other tools in the literature. For example, SEMplMe [9] uses motif analysis to predict the effects of methylation patterns measured in a BS experiment within binding site regions derived from a ChIP-seq experiment. Also, MotifMaker [10] identifies motifs associated with SMRT output methylation sequences. Here, we propose a strategy aimed to identify DNA motifs over-represented in a set of aberrantly methylated DNA sequences, obtained from a microarray-based DNA methylation experiment. This allows the user to pinpoint DNA motifs which are prone to undergo aberrant methylation in a certain biological condition. Moreover, our tool predicts transcription factor binding sites (TFBS) encompassing the identified motifs hence facilitating the exploration of possible consequences of specific DNA methylation patterns on the downstream gene transcriptional regulation. In this way, our tool supports the user in the biological interpretation of the results of a DNA methylation experiment. We developed CpGmotifs, an interactive tool based on a Graphical User Interface (GUI) implemented in R-Shiny. CpGmotifs provides a seamless and straightforward analytic flow to researchers investigating DNA motifs associated with methylation changes.

Implementation

Input

The input is constituted by one or more foreground sets of CpGs of interest (TargetCpGs) and one background set (BackCpGs). Each TargetCpG set is provided as a tab separated file containing 3 mandatory columns and an optional column. Each TargetCpG set consists of a table where each CpG is annotated with its methylation status (hyper-methylated or hypo-methylated). The background set represents the universe, typically defined as the whole set of tested CpGs, from which the TargetCpG set is derived. A flanking region size around each CpG site is used to retrieve the flanking sequences of each CpG site in the reference genome. The choice of the region size depends on various factors that are related to (1) the particular biological question, (2) the chosen motif search algorithm, and (3) the computational resources. In the next step, the tool retrieves the target and background flanking sequence sets using the BSgenome R package (<https://www.bioconductor.org/packages/release/bioc/html/BSgenome.html>).

Motif discovery

The target and the background sequence sets are then provided as the input to a de novo DNA motif search tool. For this task, we make use of the DREME [11] tool from the MEME suite, which provides in output enriched DNA motifs. Once the enrichment analysis is completed, statistically significant motifs, along with corresponding

information, are collected. In particular, for each significant motif, the following information is retrieved by parsing the DREME output xml file in R: (1) the motif consensus sequence; (2) the motif enrichment e-value; (3) the number of sequences carrying the motif in the TargetCpGs-seqs set; (4) the number of sequences not carrying the motif in the TargetCpGs-seqs set. For each discovered motif, the tool retrieves the corresponding list of CpGs containing the motif in their flanking sequence. This is accomplished by retrieving from the TargetCpGs sequences set all the motif consensus sequences reported in the DREME output. Once this set is obtained, the tool annotates the discovered motifs with the following additional fields (as reported in Additional file 1: Table S1):

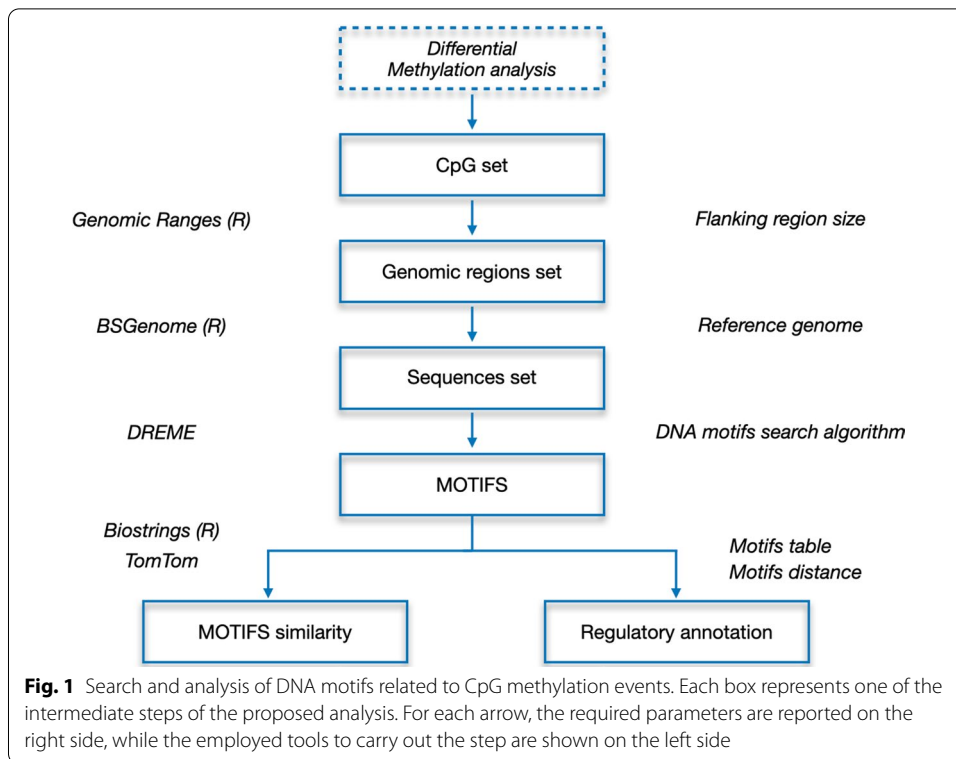
- The number of hyper-methylated CpGs in the supporting set;
- The number of hypo-methylated CpGs in the supporting set;
- The methylation ratio of the supporting set (number of hyper-methylated CpGs/number of hypomethylated CpGs);
- The methylation unbalance p-value, computed by using an enrichment test over the contingency matrix built with hyper(hypo)-methylated CpGs in the supporting set and hyper(hypo)-methylated CpGs in the background set.

Motif characterization

Once the annotated motifs are retrieved, it is possible to compare the obtained motifs among different experimental conditions or the different methylation trends, in order to get an insight on the specificity of the phenomenon under consideration. By using the *pairwiseAlignment* function from the Biostrings R package (<https://bioconductor.org/packages/release/bioc/html/Biostrings.html>), the tool computes the distance between two motifs using their consensus sequence (provided as IUPAC codes) and a nucleotide substitution matrix obtained through the *nucleotideSubstitutionMatrix* function from the Biostrings R package. The obtained similarity scores are normalized in the interval [0–1] using min-max normalization and the corresponding distance is computed as its complement to 1. Once the distance between all pairs of motifs has been obtained, a distance matrix is computed and used to perform clustering analysis. Hierarchical clustering is applied and the results are represented by means of a dendrogram adorned with methylation information from each motif by using the functions provided by the *den-dextend* R package [12].

Regulatory annotation

DNA motifs are primarily used to search for affinities between binding sites of molecules interacting with the DNA, including transcription factors (TFs). In this setting, they can be used to study which TFs show affinity with the discovered motifs, and how their function is modulated by the methylation status of the associated CpG. CpG motifs interrogates public databases of known motifs such as JASPAR [13], using the Tomtom [14] tool from the MEME suite and the obtained motifs as input. A list of highly similar motifs along with significance p-value and the corresponding binding proteins is returned and associated with each discovered motif (Fig. 1).



CpGmotifs tool

The CpGmotifs tool is built as a user-friendly and intuitive interface in order to allow the users to have a smooth experience throughout the entire analysis. A typical CpGmotifs analysis goes through three main steps, performed by using the three tabs located on the left side of the main interface. The first step consists of loading the required files and annotations, as shown in Fig. 2 (1). In particular, the user first selects the proper annotation (2) that is needed to associate genomic coordinates (and associated sequences) to CpG IDs. The user can then choose between the Illumina Infinium Human Methylation 450K (https://support.illumina.com/array/array_kits/infinium_humanmethylation_n450_beadchip_kit.html) and the Infinium MethylationEPIC (<https://www.illumina.com/products/by-type/microarray-kits/infinium-methylation-epic.html>) annotations or, alternatively, load a custom hg19 annotation. Next, the user uploads a CpG background file (3) containing all the CpG sites representing the enrichment background, including their genomic location. Similarly, the user provides one or more target CpG files (4) containing the CpGs of interest for the motif enrichment analysis. All of the uploaded files must be in tab separated format. Finally, the user may set the size of the CpG flanking region where the tool will search for enriched motifs (5).

The second step of the analysis consists in the detection of enriched CpG motifs, as shown in Fig. 3. The user can perform the analysis by selecting the “Motif analysis” tab (6), where, before starting the motifs search (7), the user has the possibility to set several parameters in order to filter and annotate the results of the motif analysis. In particular, the “motif p-value threshold” field (8) indicates the p-value of the Fisher’s Exact test for the enrichment of the positive motifs over the background sequences. The “motifs

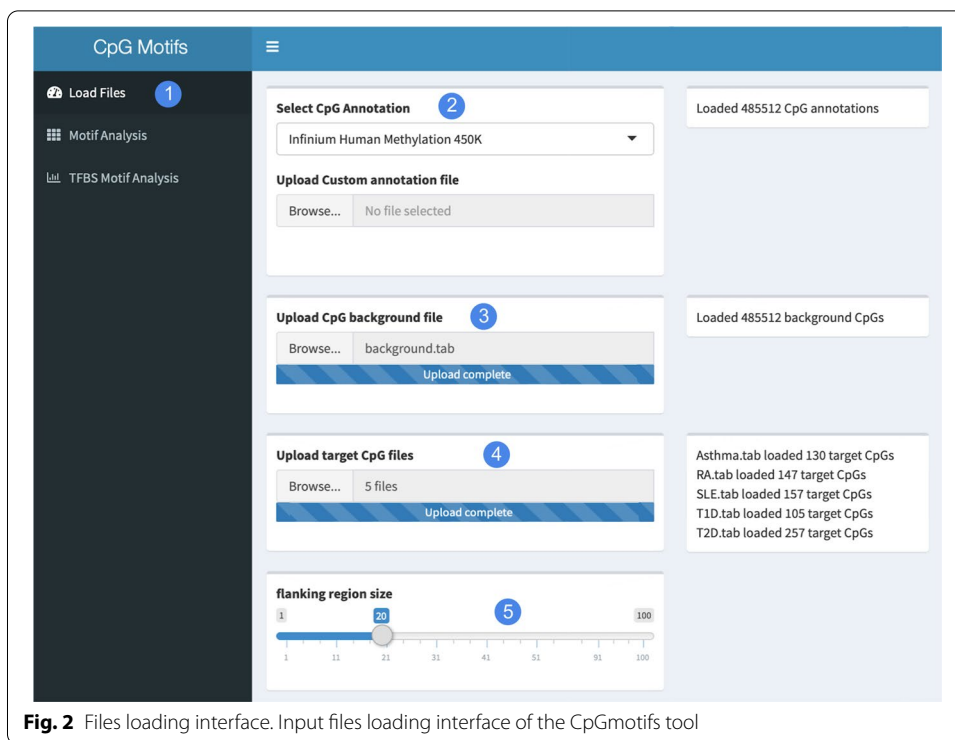


Fig. 2 Files loading interface. Input files loading interface of the CpGmotifs tool

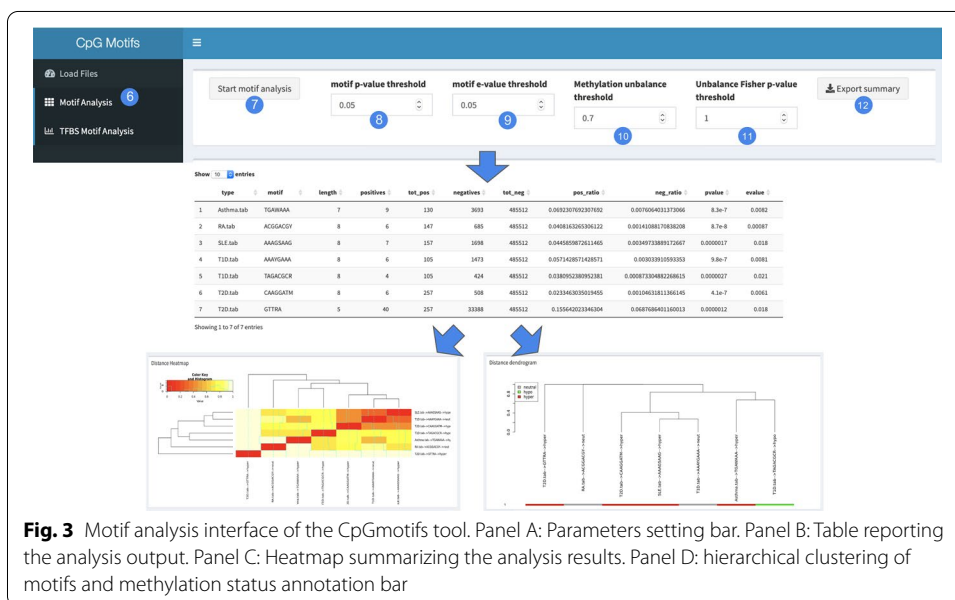


Fig. 3 Motif analysis interface of the CpGmotifs tool. Panel A: Parameters setting bar. Panel B: Table reporting the analysis output. Panel C: Heatmap summarizing the analysis results. Panel D: hierarchical clustering of motifs and methylation status annotation bar

e-value threshold” field (9) indicates the motif *p*-value times the number of candidate motifs tested. The “methylation unbalance threshold” field (10) indicates the percentage of methylated or unmethylated CpG in order to consider a certain motif as methylated, unmethylated or neutral. The “Unbalance Fisher *p*-value threshold” refers to the Fisher’s test *p*-value for the enrichment of methylated/unmethylated CpGs in the set of CpG sites carrying the motif. As a result, the tool provides a table reporting all the

enriched motifs, including their sequence, length, methylation trend and all the information regarding the statistical tests. By pressing the “Export Summary” button (12) the user can export the results table in .csv format. Along with the table, the tool produces two graphical representations of the results through a heatmap and a clustering dendrogram showing the motifs sequence similarity and their methylation trends. A clustering example is provided in Additional file 1: Figure S1.

The third step of the analysis consists in the identification of putative TFBS overlapping with the significant motifs. Also, in this case the user has the possibility to fine tune the parameters for the statistical analysis and filter the output of the analysis. In detail, the TF *p*-value threshold field (15) indicates the *p*-value threshold of the Tomtom TF enrichment test. The TF *e*-value threshold (16) indicates the E-value of the Tomtom TF enrichment test. TF *q*-value threshold (17) indicates the Benjamini & Hochberg corrected *p*-value, while the TF overlap threshold (18) indicates the minimum overlap (in the optimal alignment) of the TFBS with a certain motif to be included in the analysis. The user has the possibility to download the summary table of the obtained results as .csv file (19). Moreover, the user is provided with a heatmap reporting the TFs for which a TFBS has been identified on the columns and the samples on the rows. A transcription factors annotation example is provided in Additional file 1: Figure S2 and Table S2.

Results and discussion

Case study

We showcase the functionalities of CpGmotifs on five sets of differentially methylated probes in isolated cells from whole blood in selected susceptibility genes for inflammatory diseases, from the catalog of published genome-wide association studies (<http://www.genome.gov/gwastudies/>) [15], and reanalysed in [16]. The included diseases were asthma, atopic dermatitis, rheumatoid arthritis, systemic lupus erythematosus, type 1 and type 2 diabetes. Through the analysis of these differentially methylated probes we sought to identify whether overrepresented DNA motifs (or motifs families) exist in the context of the deregulated probes. Moreover, we aim at identifying the overall methylation trend of the identified motifs and to infer the regulatory implications of the methylation unbalance through the TFBS prediction.

In order to perform the analysis, the Infinium Human Methylation 450K annotation was chosen. The input to the CpGmotifs tool were 5 target CpG files in tab delimited format (Additional file 1) containing the probes of interest for the five above mentioned inflammatory conditions and the CpG background file containing all the probes reported in the selected annotation. The flanking region size was set to 20. The motif analysis was performed with the following parameters:

- Motif *p*-value threshold: 0.05;
- Motif *e*-value threshold: 0.05;
- Methylation unbalance threshold: 0.7;
- Unbalance Fisher *p*-value threshold: 1;

We obtained 7 significantly enriched DNA motifs over all the 5 conditions. Of these motifs, 1 was derived from asthma, 1 from rheumatoid arthritis and 1 from systemic

lupus erythematosus, while 2 motifs derived from both type 1 and type 2 diabetes. Overall, 4 motifs were characterized as mainly hyper-methylated, 1 motif as hypo-methylated and 2 as neutral. To the best of our knowledge, none of the motifs were previously associated with the relative diseases, meaning that our results could represent an unexplored ground of investigation. The Fig. 3 shows a heatmap based on the sequence similarity of the motifs, while the clustering dendrogram underlines the methylation unbalance trend for each motif.

Subsequently, we the TFBS analysis was performed with the following parameters:

- TF *p*-value threshold: 0.05;
- TF *e*-value threshold: 10;
- TF *q*-value threshold: 1;
- TF overlap threshold: 1;

As a result, a total of eight TFBS were found to be significantly associated with the identified DNA motifs (Fig. 4). Of these TFs, three were associated with the motif TGAWAA A deriving from the asthma dataset. Three TFs were associated with the motif AAAYGA AA deriving from type 1 diabetes, while 1 TF is associated with rheumatoid arthritis and 1 from systemic lupus erythematosus. The 3 TFs identified in the asthma dataset are EOMES, NFATC2, and NFAT5 and are known to regulate gene expression during immune responses. In the immune system, Eomes can positively influence the expression of IFN γ in CD8+ T-cells [17]. NFAT5 regulates the expression of the TNF α , participating in specific aspects of host defense by upregulating genes of the TNF family and other target genes in T cells subjected to osmotic stress [18]. On the other hand, NFATc2 is a key transcription factor to the pathogenesis of allergic responses [19]. It has been demonstrated that its deficiency leads to increased airway hyperresponsiveness (AHR), both after sensitization and in the absence of exogenous allergen challenge. The 3 TFs identified in type 1 diabetes are IRF1, IRF2, IRF7. Aberrations in IRF signaling cascade can lead to increased expression of type I interferon (IFN) genes, IFN-stimulated genes (ISGs), and other pro-inflammatory cytokines/chemokines, leading to the development of numerous diseases, including ones of autoimmune origin, such as type

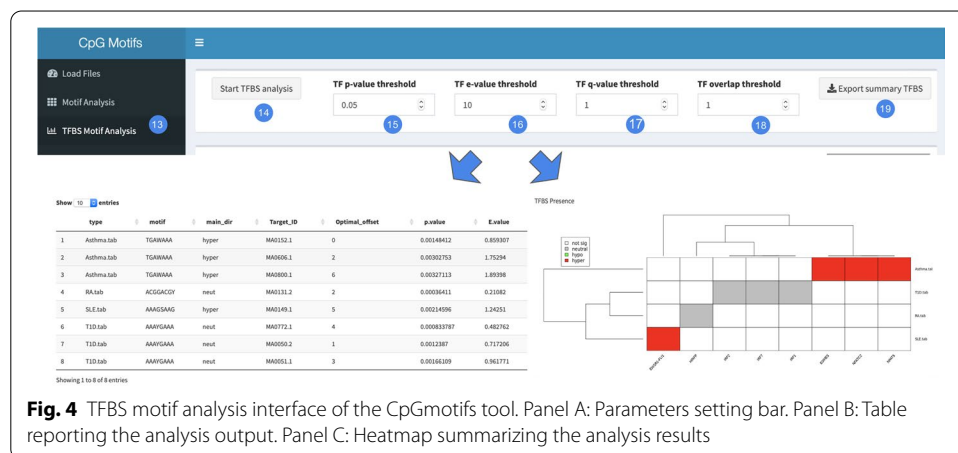


Fig. 4 TFBS motif analysis interface of the CpGmotifs tool. Panel A: Parameters setting bar. Panel B: Table reporting the analysis output. Panel C: Heatmap summarizing the analysis results

1 Diabetes. Transcription factor interferon regulatory factor 1 (IRF1), is a downstream target of IFN-gamma/signal transducer and activator of transcription (STAT)-1 and is actively involved in immune-mediated beta cell destruction. It is typically induced by IFN- γ via binding of STAT-1 to the IFN- γ -activation site in the IRF-1 promoter, but other cytokines and hormones can also trigger its expression. IRF-1 is involved in insulin secretion and, especially, in the modulation of chemokine expression by beta cells [20]. IRF2 is a negative regulator of IFN-mediated gene expression. IRF2 suppresses the activity of IRF1 by competing for binding sites. IRF7 has also been implicated in the pathogenesis of type 1 diabetes through the upregulation of inflammatory gene networks [21]. Abnormal expression of Fli-1 is important in the etiology of autoimmune diseases such as systemic lupus erythematosus and systemic sclerosis. Fli-1 is expressed in peripheral blood mononucleated cells and its overexpression correlates with severity of disease in lupus patients [22].

Conclusions

Here, we presented a novel bioinformatic procedure along with an intuitive graphical interface software to perform the analysis of the sequence context of regions related to DNA methylation events. The comprehension of the relationship linking DNA methylation alterations and proximal DNA sequences is an important step in understanding of the DNA methylation mechanisms and to better identify therapeutic targets in diseases related to methylation aberrations. The technique presented here can be applied to a wide range of public data sets, since it only requires a list of CpGs along with their methylation status. The current implementation supports CpG lists derived from relatively small (hundreds of thousands) set of tested CpGs from the human genome like data derived from the Illumina Infinium 450k or Epic platforms. However, our tool is also able to process CpG lists derived from Whole Genome or Targeted Bisulphite Sequencing. We focused the analysis on CpG methylation events, being the most studied DNA modification, but the analysis can seamlessly be applied to other cytosine modifications in the CHG and CHH contexts.

Availability and requirements

Project name: CpGmotifs

Project home page: <https://github.com/Greco-Lab/CpGmotifs>

Operating system(s): Cross Platform

Programming language: R

Other requirements: Docker

License: GPL

Any restrictions to use by non-academics: For commercial use and modifications please contact the corresponding author.

Abbreviations

AHR: Airways hyperresponsiveness; DREME: Discriminative regular expression Motif elicitation; GUI: Graphical user interface; IUPAC: International union of pure and applied chemistry; TF: Transcription factor; TFBS: Transcription factor binding site; MEME: Multiple Em for Motif Elicitation.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-021-04191-8>.

Additional file 1. Examples of output results obtained by CpGmotifs analysis.

Acknowledgements

The authors are grateful to Troy Faithfull for his critical comments on the manuscript.

Authors' contributions

GS, DG: Conceived and designed the software; GS, AF: performed the analysis and interpreted the results; GS, AF: drafted the manuscript; DG: supervised the activities and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

This study received funding from EU IMI2 project Biomap (Grant Agreement 821511). The funding body was not involved in the design of the study, collection, analysis, and interpretation of data, or in writing the manuscript.

Availability of data and materials

All the data used herein are publicly available at <http://www.genome.gov/gwastudies>.

Declarations

Ethics approval and consent to participate

The data used herein was previously published [15] and made publicly available at <http://www.genome.gov/gwastudies>.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interest.

Author details

¹Department of Biology, University of Naples "Federico II", 80126 Naples, Italy. ²Faculty of Medicine and Health Technology, Tampere University, 33014 Tampere, Finland. ³Finnish Hub for Development and Validation of Integrated Approaches (FHAIVE), Tampere University, 33014 Tampere, Finland. ⁴Institute of Biotechnology, University of Helsinki, 00014 Helsinki, Finland.

Received: 10 July 2020 Accepted: 12 May 2021

Published online: 26 May 2021

References

- Scala G, Federico A, Palumbo D, Coccozza S, Greco D. Dna sequence context as a marker of cpG methylation instability in normal and cancer tissues. *Sci Rep*. 2020. <https://doi.org/10.1038/s41598-020-58331-w>.
- Feltus FA, Lee EK, Costello JF, Plass C, Vertino PM. Predicting aberrant cpG island methylation. *Proc Natl Acad Sci*. 2003;100(21):12253–8. <https://doi.org/10.1073/pnas.2037852100>.
- Boyes J, Bird A. Repression of genes by dna methylation depends on cpG density and promoter strength: evidence for involvement of a methyl-cpG binding protein. *EMBO J*. 1992;11(1):327–33. <https://doi.org/10.1002/j.1460-2075.1992.tb05055.x>.
- Baubec T, Schbeler D. Genomic patterns and context specific interpretation of dna methylation. *Curr Opin Genet Dev*. 2014;25:85–92. <https://doi.org/10.1016/j.jgde.2013.11.015>.
- Bartlett TE, Zaikin A, Olhede SC, West J, Teschendorff AE, Widschwendter M. Corruption of the intra-gene dna methylation architecture is a hallmark of cancer. *PLoS ONE*. 2013;8(7):68285. <https://doi.org/10.1371/journal.pone.0068285>.
- Hansen KD, Timp W, Bravo HC, Sabuncuyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep DEA. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*. 2011;43(8):768–75. <https://doi.org/10.1038/ng.865>.
- Wang M, Zhang K, Ngo V, Liu C, Fan S, Whitaker JW, Chen Y, Ai R, Chen Z, Wang JEA. Identification of dna motifs that regulate dna methylation. *Nucl Acids Res*. 2019;47(13):6753–68. <https://doi.org/10.1093/nar/gkz483>.
- Sanosaka T, Imamura T, Hamazaki N, Chai M, Igarashi K, Ideta-Otsuka M, Miura F, Ito T, Fujii N, Ikeo KEA. Dna methylome analysis identifies transcription factor-based epigenomic signatures of multilineage competence in neural stem/progenitor cells. *Cell Rep*. 2017;20(12):2992–3003. <https://doi.org/10.1016/j.celrep.2017.08.086>.
- Nishizaki, S.S., Boyle, A.P.: SEMPLME: A tool for integrating dna methylation effects in transcription factor binding affinity predictions. *bioRxiv* (2021). <https://doi.org/10.1101/2020.08.13.250118>.
- Li T, Zhang X, Luo F, Wu F-X, Wang J. Multimotifmaker: a multi-thread tool for identifying dna methylation motifs from pacbio reads. *IEEE/ACM Trans Comput Biol Bioinform*. 2020;17(1):220–5. <https://doi.org/10.1109/tcbb.2018.2861399>.
- Bailey TL. Dreme: motif discovery in transcription factor chip-seq data. *Bioinformatics*. 2011;27(12):1653–9. <https://doi.org/10.1093/bioinformatics/btr261>.

12. Galili T. dendextend: an r package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*. 2015;31(22):3718–20. <https://doi.org/10.1093/bioinformatics/btv428>.
13. Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, vanderLee R, Bessy A, Chneby J, Kulkarni SR, Tan GEA. Jaspar 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucl Acids Res*. 2017;46(D1):1284–1284. <https://doi.org/10.1093/nar/gkx1188>.
14. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble W. Quantifying similarity between motifs. *Genome Biol*. 2007;8(2):24. <https://doi.org/10.1186/gb-2007-8-2-r24>.
15. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci*. 2009;106(23):9362–7. <https://doi.org/10.1073/pnas.0903103106>.
16. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahl S-E, Greco D, Sderhll C, Scheynius A, Kere J. Differential dna methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS ONE*. 2012;7(7):41361. <https://doi.org/10.1371/journal.pone.0041361>.
17. Knox JJ, Cosma GL, Betts MR, McLane LM. Characterization of t-bet and eomes in peripheral human immune cells. *Front Immunol*. 2014. <https://doi.org/10.3389/fimmu.2014.00217>.
18. Neuhofer W. Role of nfat5 in inflammatory disorders associated with osmotic stress. *Curr Genom*. 2010;11(8):584–90. <https://doi.org/10.2174/138920210793360961>.
19. Karwot R, Maxeiner JH, Schmitt S, Scholtes P, Hausding M, Lehr HA, Glimcher LH, Finotto S. Protective role of nuclear factor of activated t cells 2 in cd8+ long-lived memory t cells in an allergy model. *J Allergy Clin Immunol*. 2008;121(4):992–9996. <https://doi.org/10.1016/j.jaci.2007.12.1172>.
20. Gysemans C, Callewaert H, Moore F, Nelson-Holte M, Overbergh L, Eizirik DL, Mathieu C. Interferon regulatory factor-1 is a key transcription factor in murine beta cells under immune attack. *Diabetologia*. 2009;52(11):2374–84. <https://doi.org/10.1007/s00125-009-1514-5>.
21. Heinig M, Petretto E, Wallace C, Bottolo L, Rotival M, Lu H, Li Y, Sarwar R, Langley SR, Bauerfeind AEA. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature*. 2010;467(7314):460–4. <https://doi.org/10.1038/nature09386>.
22. Li Y, Luo H, Liu T, Zacksenhaus E, Ben-David Y. The ets transcription factor flt-1 in development, cancer and disease. *Oncogene*. 2014;34(16):2022–31. <https://doi.org/10.1038/ncr.2014.162>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

