

methGraph

A genome visualization tool for PCR-based methylation assays

Steve Lefever,^{1,†} Jasmien Hoebeek,^{1,†} Filip Pattyn,¹ Gábor Tusnády,² Anne De Paepe,¹ Frank Speleman,¹ Tamás Arányi² and Jo Vandesompele^{1,*}

¹Center for Medical Genetics; Ghent University Hospital; Ghent, Belgium; ²Institute of Enzymology; BRC; Hungarian Academy of Sciences; Budapest, Hungary

[†]These authors contributed equally to this paper.

Key words: DNA methylation, graphics tool, methylation-specific PCR (MSP), bisulfite sequencing (BiSeq), CpG island, CpG dinucleotides, primer specificity, assay visualization, UCSC genome browser

Abbreviations: BiSeq, bisulfite sequencing; CCDS, consensus coding DNA sequence; COBRA, combined bisulfite restriction analysis; ePCR, electronic PCR; MSP, methylation-specific PCR; PCR, polymerase chain reaction; SNP, single nucleotide polymorphism; PDF, portable document format; EPS, encapsulated postscript; BED, browser extensible display

Abnormalities in DNA methylation of CpG islands that play a role in gene regulation affect gene expression and hence play a role in disease, including cancer. Bisulfite-based DNA methylation analysis methods such as methylation-specific PCR (MSP) and bisulfite sequencing (BiSeq) are most commonly used to study gene-specific DNA methylation. Assessing specificity and visualizing the position of PCR primers in their genomic context is a laborious and tedious task, primarily due to the sequence changes induced during the bisulfite conversion. For this purpose, we developed methGraph, a web application for easy, fast and flexible visualization and accurate in silico quality evaluation of PCR-based methylation assays. The visualization process starts by submitting PCR primer sequences for specificity assessment and mapping on the genome using the BiSearch ePCR primer-search algorithm. The next step comprises the selection of relevant UCSC genome annotation tracks for display in the final graph. A custom track showing all individual CpG dinucleotides, representing their distribution in the CpG island is also provided. Finally, methGraph creates a BED file that is automatically uploaded to the UCSC genome browser, after which the resulting image files are extracted and made available for visualization and download. The generated high-quality figures can easily be customized and exported for use in publications or presentations. methGraph is available at <http://mellfire.ugent.be/methgraph/>.

Introduction

DNA hypermethylation is frequently observed at promoter regions of both coding and non-coding genes and induces silencing of these genes. This naturally occurring mechanism is involved in X-inactivation, and genetic imprinting, and occurs during mammalian development and differentiation. However, in cancer cells abnormal CpG island promoter methylation may cause transcriptional repression of tumor suppressor genes.¹ PCR-based methylation assays such as methylation-specific polymerase chain reaction (MSP) and bisulfite sequencing (BiSeq) have become the gold standards over the last few years to gain insights in the patterns of aberrant methylation.^{2,3} These methods make use of bisulfite converted DNA whereby unmethylated cytosines are deaminated to uracils, while 5-methylcytosines are protected and remain unchanged. Primer similarity search tools, such as methBLAST

and the BiSearch electronic PCR (ePCR) tool, have been developed to check the specificity of the primers on the altered genome template.⁴⁻⁶ However, additional in silico evaluation of designed primers (absence of single nucleotide polymorphisms (SNPs) in the primer annealing region, genomic localization and context,...) remains time consuming due to the induced sequence changes and the non-complementarity of the sense and antisense DNA strand as a result of the bisulfite conversion. Current primer design tools (e.g., MethPrimer, Methyl Primer Express and MSPprimer) offer a schematic representation of the input sequence, the CpG island and the designed primers.^{7,8} However, they often lack specificity assessment and the visual output does not provide useful genomic information nor the flexibility to manipulate the generated figure in order to select the most appropriate primer pair for a given methylation assay. Recently, the advanced MethMarker tool has been developed.⁹ Besides its use as a primer design tool

*Correspondence to: Jo Vandesompele; Email: Joke.Vandesompele@UGent.be

Submitted: 11/13/09; Accepted: 01/11/10

Previously published online: www.landesbioscience.com/journals/epigenetics/article/11161

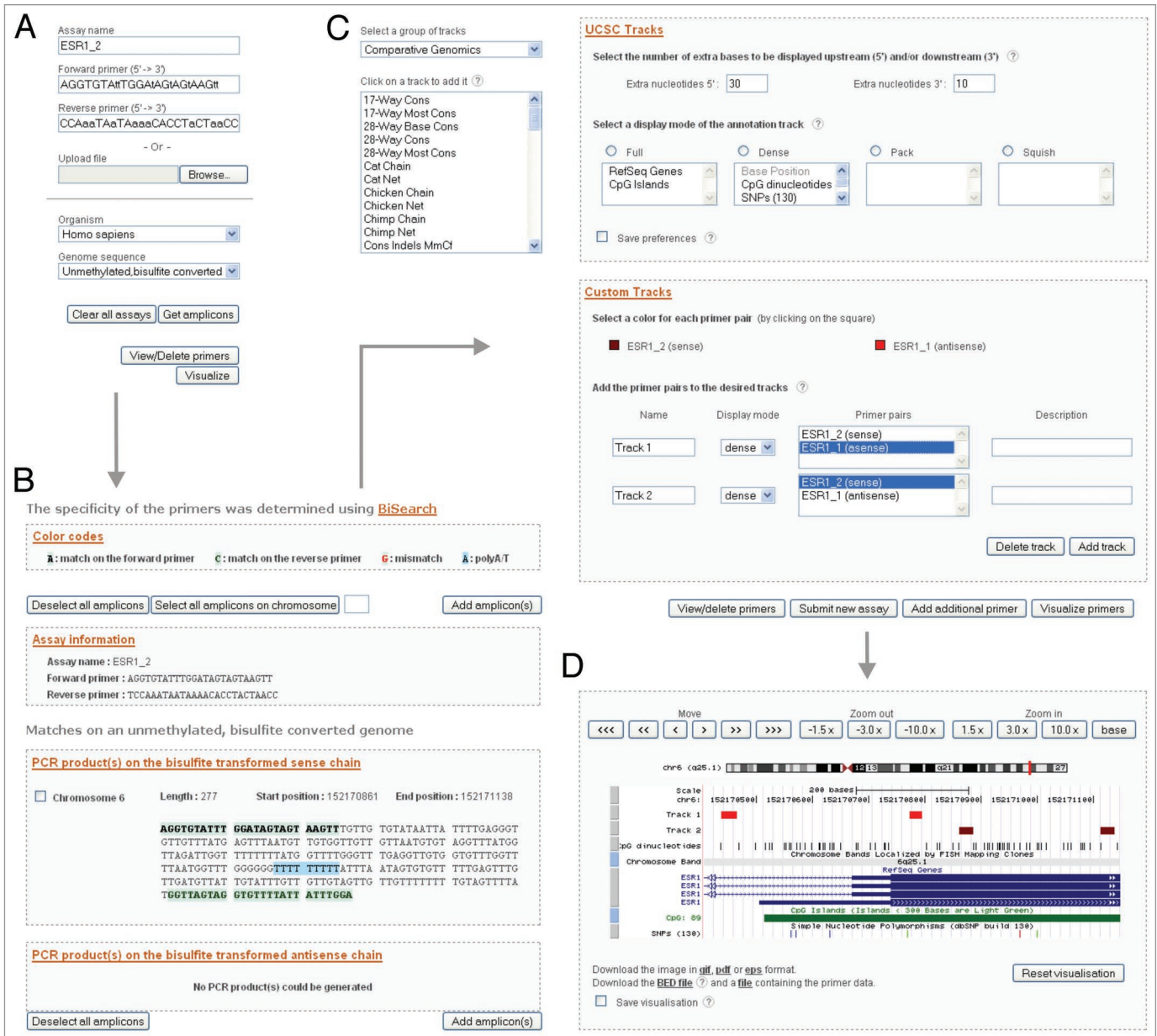


Figure 1. Flowchart demonstrating the different steps in using methGraph. (A) Primer sequences are entered one by one or in batch followed by selecting the target organism and genome sequence for primer specificity assessment. (B) The specificity of the primer pairs is determined using the BiSearch ePCR algorithm. Specific amplicons are selected by the user for visualization. Selected primer pairs can be viewed or removed if needed. (C) Specifying a color for each primer pair and selection of extra nucleotides up or downstream of the viewed region is optional. The visualization of the primer pairs and additional information is customizable by selecting UCSC tracks or creation of custom tracks. (D) The resulting figure contains all selected UCSC and custom tracks. It is possible to zoom and to move left and right over the figure.

it comprises a bioinformatic workflow to translate known differentially methylated regions into clinical biomarkers. Currently, however, no tool is available to visualize primers developed for different types of assays, e.g., BiSeq versus MSP, in order to be able to validate the results obtained by one method with the other. To overcome all these limitations, we designed methGraph, a powerful tool in epigenetic research for accurate evaluation and visualization of PCR-based methylation assays. MethGraph is available at <http://mellfire.ugent.be/methgraph>.

Results

MethGraph is a web application that combines the ePCR primer specificity assessment of BiSearch (<http://bisearch.enzim.hu/>) with the UCSC Genome Browser (<http://genome.ucsc.edu/>),^{10,11} Primer validation and genomic context assessment are thus integrated into a single tool. MethGraph is easy to use and consists of three parts (Fig. 1). The first step comprises the submission of one or more primer pairs either separately (forward and reverse primer sequences from the 5' to the 3' end) or by means of a plain

text file containing a list of primer pairs. In a second step the organism of interest (*Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Pan troglodytes* (others on request)) and the type of genome sequence (original unconverted, unmethylated bisulfite converted and methylated bisulfite converted) for which the primers' specificity must be checked, need to be selected. Analysis of the primer specificity is performed by the BiSearch ePCR algorithm which presents possible amplicons in order of decreasing expectancy. Primer mismatches, possibly representing non-specific amplicons, are indicated in red, while poly-A and poly-T stretches are marked in blue. Amplicons of interest can be selected for visualization and additional primer pairs can be added if desired. Note that if the submitted primer pairs anneal to sequences located on different chromosomes, one can visualize the primers that yield a predicted amplicon on the chromosome of choice. After target amplicon selection, default UCSC tracks and a custom made CpG dinucleotide track, visualizing all the individual CpG dinucleotides, can be selected and represented in different UCSC browser display modes (dense, squish, pack and full). The next step comprises the formatting (display mode, color, name and description) and visualization of the primer custom tracks. Once an image is generated, native UCSC functionality is available, such as moving along the genome, zooming in or out, or obtaining additional information concerning tracks (by clicking on the track when displayed in full mode). All genomic features available in the UCSC genome browser can be combined with an unlimited number of pre-designed primers on three types of genome sequences (i.e., original unconverted, unmethylated bisulfite converted and methylated bisulfite converted), from four different species. Finally, methGraph provides the opportunity for users to save the visualization for future use (saved visualizations can be retrieved under the 'Your visualizations' menu) or to download the generated output in a data file (containing a summary of the primer data) or a BED file for further processing in a genome browser of interest. Two additional export formats are available (PDF or EPS) making methGraph the ideal tool for creating customizable high quality graphics for research papers, posters and presentations. An illustrated manual covering all steps explained above is available for download in PDF format at <http://mellfire.ugent.be/methgraph/manual.php>.

Discussion

Although many methylation-related bio-informatics tools are currently available, none of them provide an integrated portal for quality assessment nor generates informative and attractive figures of PCR-based methylation assays in a straightforward and quick way. With methGraph, we provide a flexible and powerful tool that combines the in silico assay evaluation and the visualization of such assays in their genomic context using UCSC tracks.

Evaluation of primer specificity of bisulfite-based methylation assays is of great importance to prevent amplification of non-specific sequences due to a reduced sequence complexity after bisulfite conversion. To check the primer specificity, methGraph uses the BiSearch sequence alignment tool to map primer sequences on the genome. BiSearch queries on the different genome sequences

are performed faster compared to the methBLAST algorithm, thanks to the use of an innovative hashing technique. Another aspect of primer quality is to ensure the absence of SNPs residing within the annealing site. If SNPs are present, primer annealing can be compromised resulting in suboptimal sensitivity or lack of amplification of the polymorphic fragment. By selecting the UCSC SNP track, SNPs will be displayed in the generated figure with a link to NCBI's dbSNP when displayed in full mode to retrieve additional information.

One of the major advantages of methGraph is the possibility to visualize different types of PCR-based methylation assays (e.g., MSP, BiSeq, methylation-sensitive high-resolution melting analysis) in one view. As such, primers from different methylation assays can be compared on the level of localization, CpG dinucleotide coverage, overlap between primer pairs, genomic context, etc. In many studies, the methylation status of the promoter region of a gene of interest is first assessed by BiSeq to reveal the CpG dinucleotides with a methylation status representative for the inactivation of the gene. Further on, these CpG sites are analyzed on a larger set of samples by MSP. This is mainly done because MSP is more cost-effective and more straightforward than BiSeq.

Furthermore, methGraph can be used to evaluate and visualize primers found in literature and databases such as MethPrimerDB.⁵ A direct link to methGraph is provided in MethPrimerDB making it possible to check whether these primers meet the quality standards as described above. In the future, we plan the integration of methGraph in a primer design pipeline, making the design and selection of primers for a certain assay more straightforward and with a higher success rate.

Although methGraph was originally designed for methylation-specific assays, it can also be used for conventional PCR assays on the original unconverted normal genome sequence. Depending on which UCSC tracks are selected for visualization, different questions can be answered. Examples include (1) the quick evaluation of the occurrence of SNPs in primer binding sites and (2) the assessment of whether sequencing primers cover the entire coding region of the gene of interest by selecting the consensus coding DNA sequence (CCDS) gene annotation track. This aspect makes methGraph a multi functional tool that can also be used by researchers outside the field of DNA methylation (Fig. 2).

Finally, the produced images are editable as vector graphics making them perfectly suitable as high quality figures in presentations, posters and publications.

In brief, methGraph is an indispensable attribute in the expanding field of PCR based methylation research, having the potential to both save time and provide unparalleled visual output.

Materials and Methods

methGraph is developed as a web-based application using the scripting languages PHP (PHP Hypertext Preprocessor—v5.1.6) for server-side data-processing and database connectivity and custom-made javascript code for client-side effects in the interface. User data and saved visualizations are stored in a MySQL

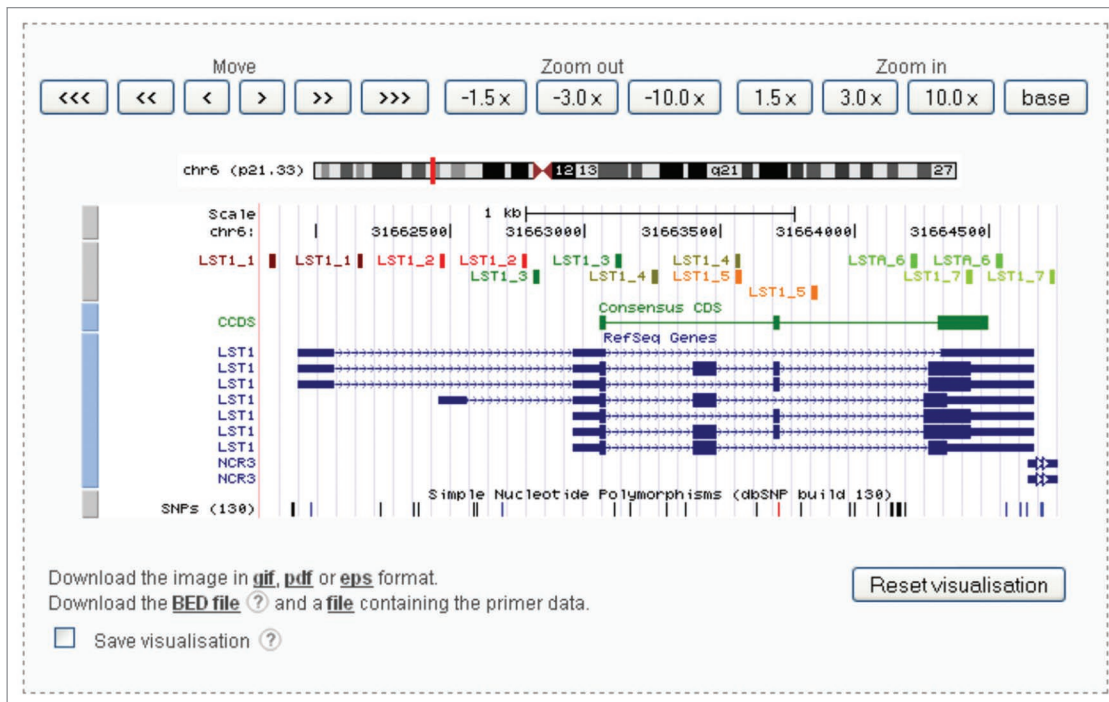


Figure 2. Applicability of methGraph outside the methylation field, e.g., mapping of sequencing primer sets that cover the coding region of an entire gene. methGraph can be used to visualize primers used in several types of assays. Here, we have designed primer pairs for amplification and sequencing of all exons of *LST1*. This visualization allows to check the primer positions relative to SNPs which could influence the amplification efficiency of the primers, and nicely shows that all exons of the different splice variants are targeted.

relational database (v 14.12, distribution 5.0.77) and accessed through PHP scripting.

During the processing of the submitted primer assay, methGraph sends the required data to a local BiSearch server in the background to map the primers on the chosen organism and genome type (original unconverted, unmethylated bisulfite converted and methylated bisulfite converted). BiSearch outputs the generated amplicons together with mapping data. And the user can select the desired amplicons and choose additional information tracks to display and start the visualization process.

To visualize the submitted data, methGraph makes use of the online UCSC Genome Browser. Locally, a BED file is created containing the mapping information of the primers and PCR amplicons as a custom track and the selected UCSC tracks. Detailed information about the BED file format is available in the FAQ section of the UCSC genome browser.^{10,11} Further on, the BED file is uploaded to the UCSC Genome Browser by means of libcurl (v7.15.5) for processing. The resulting session ID from the genome browser is send back to methGraph to display the results in the interface.

References

1. Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum Mol Genet* 2007; 16:50-9.
2. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci USA* 1992; 89:1827-31.
3. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci USA* 1996; 93:9821-6.
4. Aranyi T, Tusnady GE. BiSearch: ePCR tool for native or bisulfite-treated genomic template. *Methods Mol Biol* 2007; 402:385-402.
5. Pattyn F, Hoebeeck J, Robbrecht P, Michels E, De Paep A, Bottu G, et al. methBLAST and methPrimerDB: web-tools for PCR based methylation analysis. *BMC Bioinformatics* 2006; 7:496.
6. Tusnady GE, Simon I, Varadi A, Aranyi T. BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Res* 2005; 33:9.
7. Brandes JC, Carraway H, Herman JG. Optimal primer design using the novel primer design program: MSPprimer provides accurate methylation analysis of the ATM promoter. *Oncogene* 2007; 26:6229-37.
8. Li LC, Dahiya R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 2002; 18:1427-31.

Acknowledgements

We would like to thank Justine Nuytens for evaluating the tool and for her useful comments and suggestions for improving it. S.L. is supported by a Ghent University grant (BOF 01Z09407). J.H. and F.P. are postdoctoral researchers supported by a grant of the Ghent University (respectively BOF 01P07406 and BOF 01F07207) and by the Fund for Scientific Research Flanders (Krediet aan Navorsers 1.5.207.08 and 1.5.118.09.N.00, respectively). Tamás Arányi has a Hungarian Research Fund (OTKA) grant PD79183, Gábor E. Tusnady is supported by a Bolyai Janos scholarship and has an OTKA grant K75460. This article presents research results of the Belgian program of Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. This study was supported by the Children Cancer Fund Ghent, the "Stichting tegen Kanker", FWO-Grant G.0198.08, the Methusalem program BOF08/01M01108. We acknowledge the support of the European Community under the FP6 (project: STREP: EET-pipeline, number: 037260).

-
9. Schüffler P, Mikeska T, Waha A, Lengauer T, Bock C. MethMarker: user-friendly design and optimization of gene-specific DNA methylation assays. *Genome Biol* 2009; 10:105. Epub 2009.
 10. Kuhn RM, Karolchik D, Zweig AS, Wang T, Smith KE, Rosenbloom KR, et al. The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res* 2009; 37:755-61.
 11. Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. UCSC genome browser tutorial. *Genomics* 2008; 92:75-84.

©2010 Landes Bioscience.
Do not distribute.