Automated Methylation Sequence Analysis with Mutation Surveyor®

David Hulce, ChangSheng, Jonathan Liu, Softgenetics LLC

Introduction

Methylation of DNA at the C of CpG sites has been associated with X chromosome inactivation, genomic imprinting, embryonic development, human diseases[1]; and differential gene expression[2]. Epigenetic mechanisms involving methylation of DNA have been reported in colon cancer[3]. Differential DNA methylation can be used to identify fetal DNA in maternal blood[4]. Bisulfite treatment of DNA followed by sequencing is one method to analyze methylation of DNA[5]. Improving speed and efficiency of sequence trace analysis will further the understanding of DNA methylation and its relationship to biological processes.

Bisulfite treatment of single-stranded DNA converts cytosine to uracil; methylated cytosine is not converted. Nucleotide sequencing of treated DNA will show unmethylated cytosines as thymines; 5-methylcytosines will be read as cytosines. Sequencing electropherograms from bisulfite-treated DNA, when compared to a reference sequence, that show Cs of CpGs as TpGs can be interpreted as these Cs of these CpGs are not methylated. If this comparison shows CpGs as unchanged, the C of these CpGs can be interpreted as being methylated.

Mutation Surveyor is a powerful software using a unique physical trace comparison technology to detect nucleotide changes between two sequence traces. Mutation Surveyor has a new function to simplify the analysis of DNA methylation using bisulfite conversion of cytosine to uracil followed by nucleotide sequencing.

Methylation Sequence Data Analysis
The GenBank sequence of the DNA of interest is used as a "ruler" to report nucleotide changes, including methylations and mutations. The methylation function of Mutation Surveyor utilizes the GenBank sequence text or derivatives of the sequence text and converts it into a synthetic reference sequence. This synthetic reference sequence is then physically compared to the sample sequence traces from bisulfite-treated DNA to find nucleotide differences.

Procedure

1. Set Mutation Surveyor software to the Methylation application:
In the menu bar, Process-Settings-Others-Methylation.
If Methylation is **NOT** checked, the software will process the data as normal mutation detection. The user can choose one of two methods to modify the GenBank sequence: Set by User or Auto Methylation. We recommend using Auto Methylation. Mutation Surveyor will choose the best conditions so that the number of mutations detected is minimized, maintaining accuracy and simplifying data analysis.

Figure 1: Methylation Options: Auto Methylation will modify the GenBank sequence to give the synthetic reference trace with minimum number of nucleotide changes. User may also set the sequence modification.

2. Open Files to be Analyzed:
a. You **must** input the GenBank sequence file into the GenBank Sequence File (Optional) box. The GenBank file **must** be in the forward strand orientation so that CpG is well defined. The **reverse complement** sequence should be converted to forward strand using Tools'- Advanced GBK File Editor and saved as reverse complement.
b. The reference trace is synthesized from the derivative of the GenBank file(s).
Please do **NOT** input any reference files into the box labeled Reference Files.
c. Input the sample trace files into the box labeled Sample Files.

3. Run the Analysis: In the menubar, select Process-Run

4. User Interface and Data Interpretation: Peak-by-peak comparison of the derivative and sample traces (Fig.2) can be viewed in the Graphic Analysis Display window. Possible methylation sites will be identified in the mutation electropherogram.

Figure 2: Graphic Analysis Display: The top panel is the reference trace synthesized from the derivative GenBank file. The blue lines above the text indicate that the GenBank text was C (blue color). The second panel is the sample trace. The third panel is the mutation detection panel.
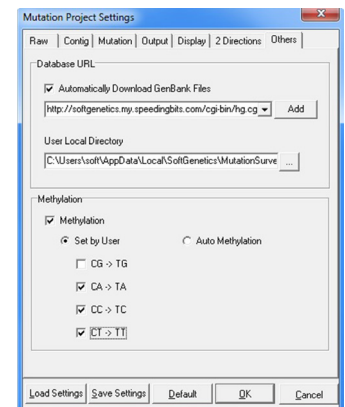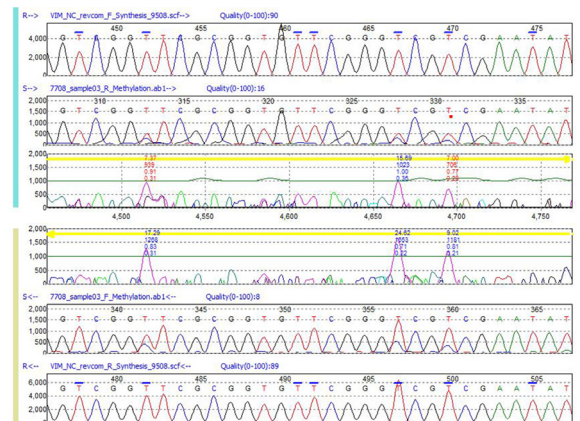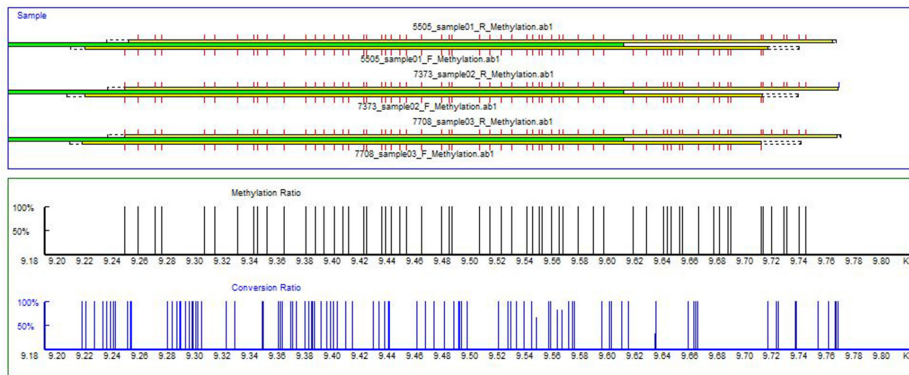

Figure 1

Figure 2

5. Data Analysis and Output: A methylation report (Fig. 3) can be obtained by clicking on the icon for the Customer Output and selecting the template—Methylation—in the Output Settings window. Click on the disk icon to save the report as a .txt file for printing.

				7373_sampl	7373_sampl	7708_sampl	7708_sampl	5505_sampl	5505_sampl
9349	C	c	M	M	M	M	M	M	M
9350	G	g	·	·	·	·	·	·	·
9351	T	t	·	·	·	·	·	·	·
9352	C	c	M	M	M	M	M	M	M
9353	G	g	·	·	·	·	·	·	·
9354	G	g	·	·	·	·	·	·	·
9355	C	t	S	S	S	S	S	S	S
9356	C	t	S	S	S	S	S	S	S
9357	A	a	·	·	·	·	·	·	·
9358	G	g	·	·	·	·	·	·	·
9359	C	c	M	M	M	M	M	M	M
9360	G	g	·	·	·	·	·	·	·
9361	A	a	·	·	·	·	·	·	·
9362	G	g	·	·	·	·	·	·	·

Figure 3: Methylation Report Table: M represents methylated and U represents unmethylated. S stands for successful conversion of C to T and I represents incomplete conversion.

6. Graphical Analysis Overview: Clicking the Graphic Display icon in the Methylation Report window, (figure 3) will show the pattern of methylated- and unmethylated-Cs along the DNA fragments used for analysis. Cs not part of a CpG dinuclotide will not be included.

Figure 4: Graphic Output of Methylation Results: Red color tics represent methylated sites and blue color tics represent unmethylated locations.

Results

Figure 2 shows three methylation sites: one at frame 4515 in the sequence G-C-T, one at frame 4665 in the sequence G-C-C, and one at frame 4695 in the sequence G-C-C (compare sequences G-T-T, G-T-C, and G-T-C of synthetic reference trace to sample trace). The mutation electropherogram identifies these T to T/C mutations as possible methylation sites.

The methylation report in figure 3 shows that the C at nucleotide position 9352 is methylated in all three samples. Cytosines that are not contained in a dinucleotide sequence CpG will be converted to T. The C's at positions 9355 and 9356 have been recorded as S: successfully converted from C to T. These C's are not located in a CpG dinucleotide sequence, and therefore, they are not recorded as unmethylated.

Discussion

There are a variety of techniques to study methylation of DNA: methylation-sensitive restriction endonuclease digestion/Southern blotting, methylated CpG island amplification (MCA), sequencing bisulfite-converted DNA and methylation sensitive PCR (MSP), to name some. The new function of methylation detection in sequencing bisulfite-converted DNA quickly and efficiently identifies nucleotide position and methylation status of Cs in multiple DNA samples. The methylation output table distinguishes between Cs contained in CpG dinucleotides (U, unmethylated; M, methylated) and Cs contained in other sequences (S or I). Other mutations that are detected will be recorded in the methylation report.

Methylation status of cytosine in CpG dinucleotides and its effect on development, gene expression and disease is an interesting field of study. Mutation Surveyor is an accurate, user-friendly software tool which can simplify analysis of sequencing traces generated in DNA methylation studies.

SOFTGENETICS®
Software PowerTools for Genetic Analysis

SoftGenetics LLC 100 Oakwood Ave. Suite 350 State College, PA 16803 USA
Phone: 814/237/9340 Fax 814/237/9343
www.softgenetics.com email: info@softgenetics.com

MUTATION **Surveyor®**
DNA Variant Analysis Software

Results

1. DNA Methylation and Human Disease. Keith D. Robertson. Nature Reviews Genetics 6:597-610 (2005).
2. Association of Tissue-specific Differentially Methylated Regions (TDMs) with Differential Gene Expression. Fei Song, Joseph F. Smith, Makoto T. Kimura, Arlene D. Morrow, Tomoki Matsuyama, Hiroki Nagase and William A. Held. Proc. Natl. Acad. Sci., USA. 102:3336-3341 (2005).
3. Aberrantly Methylated CDKN2A, MGMT, and MLH1 in Colon Polyps and in Fecal DNA from Patients with Colorectal Polyps. Zsolt Petko, Mahan Ghiassi, Anthony Shuber, Janice Gorham, Walter Smalley, M. Kay Washington, Stephen Schultenover, Shiva Gautam, Sanford D. Markowitz and William M. Grady. Clinical Cancer Research 11:1203-1209 (2005).
4. Detection of the Placental Epigenetic Signature of the Maspin Gene in Maternal Plasma. Stephen S. C. Chim, Yu K. Tong, Rossa W. K. Chiu, Tze K. Lau, Tse N. Leung, Lisa Y. S. Chan, Cees B. M. Oudejans, Chunming Ding, and Y. M. Dennis Lo. Proc. Natl. Acad. Sci., USA. 102:14753-14758 (2005).
5. A Genomic Sequencing Protocol that Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands. Frommer, M., L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy and C. L. Paul. Proc. Natl. Acad. Sci., USA. 89:1827-1831 (1992).

Acknowledgements

SOFTGENETICS®
Software PowerTools for Genetic Analysis

SoftGenetics LLC 100 Oakwood Ave. Suite 350 State College, PA 16803 USA
Phone: 814/237/9340 Fax 814/237/9343
www.softgenetics.com email: info@softgenetics.com

MUTATION
Surveyor®
DNA Variant Analysis Software