Quantification of gene-specific methylation of *DNMT3B* and *MTHFR* using sequenom EpiTYPER®

Vikki Ho a,*, Janet E. Ashbury a, Sherryl Taylor b, Stephen Vanner c, Will D. King a

a Department of Public Health Sciences, Queen’s University Kingston, Ontario, Canada K7L3N6  
b Department of Medical Genetics, University of Alberta, Edmonton, Alberta, Canada T6G 2R3  
c Gastrointestinal Diseases Research Unit (GIDRU), Queen’s University, Kingston, Ontario, Canada K7L3N6

**A R T I C L E   I N F O**

Article history:  
Received 29 September 2015  
Received in revised form  
9 November 2015  
Accepted 16 November 2015  
Available online 26 November 2015

**A B S T R A C T**

Among 272 patients undergoing a screening colonoscopy, DNA methylation of *DNMT3B* and *MTHFR*, genes encoding enzymes critical to one-carbon metabolism, was quantified in blood leukocytes using Sequenom EpiTYPER®. DNA methylation was quantified in 66 and 28 CpG sites of *DNMT3B* and *MTHFR* respectively, and conceptualized using two approaches. First, measures representing average methylation across all CpG sites were created. Second, unsupervised principal component (PC) analysis was used as a pattern derivation and data-reduction approach, to develop two summary variables (PC1 and PC2). These two summary variables represented methylation around the transcription start site (PC1) and in the gene-coding area (PC2) for both *DNMT3B* and *MTHFR*. The data contained in this article presents the variation of methylation levels for individual CpG sites within the *DNMT3B* and *MTHFR* genes and possible correlations uncovered using PC analysis. The data are related to the research article “Gene-specific DNA methylation of DNMT3B and MTHFR and colorectal adenoma risk” in Mutation Research – Fundamental and Molecular Mechanisms of Mutagenesis.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Specifications table

| Subject area                      | Epidemiology                              |
|-----------------------------------|-------------------------------------------|
| More specific subject area        | Cancer, Epigenetics                        |
| Type of data                      | Table, figure                             |
| How data was acquired             | Sequenom EpiTYPER                         |
| Data format                       | Raw data                                  |
| Experimental factors              | Leukocytic DNA was isolated from blood samples collected from 272 patients undergoing a screening colonoscopy. Ninety participants were diagnosed with one or more pathologically-confirmed tubular, tubulo-villous or villous adenoma(s), while 182 patients had no abnormality identified during colonoscopy. |
| Experimental features             | The Sequenom EpiTYPER® technology was used to quantify DNA methylation in 66 and 28 CpG sites of DNMT3B and MTHFR. |
| Data source location              | Kingston, Ontario, Canada                  |
| Data accessibility                | Data is available with this article        |

Value of the data

- Using the Sequenom EpiTYPER® technology, we quantified methylation levels of 66 and 28 CpG sites located within the promoter regions of DNMT3B and MTHFR, respectively; methylation levels of these CpG sites were highly variable. Future etiological studies investigating the role of genespecific DNA methylation in DNMT3B and MTHFR could explore different conceptualizations of methylation including principal component analysis.
- The methodology used in processing raw data values for methylation levels quantified using the Sequenom EpiTYPER® technology could be used as a benchmark for future research.
- The data derived from our study can be used to inform future investigations relating methylation levels within specific CpG sites to gene expression for DNMT3B and MTHFR.

1. Data

DNA methyltransferase 3B (DNMT3B: NM_006892.3) and methylenetetrahydrofolate reductase (MTHFR; NM_005957.4) are two key genes which encode enzymes critical to 1C metabolism, disruption of which has increasingly been implicated in colorectal cancer etiology [1,2]. DNA methylation involves the covalent addition of a methyl group to cytosine residues that precede guanine (CpG) and is the most commonly studied epigenetic process [3]. In the present work, among 272 patients undergoing a screening colonoscopy, we present the data of leukocytic methylation levels quantified using the Sequenom EpiTYPER® technology for 66 and 28 CpG sites in the DNMT3B and MTHFR genes, respectively. The data are related to the research article “Gene-specific DNA methylation of DNMT3B and MTHFR and colorectal adenoma risk” in Mutation Research – Fundamental and Molecular Mechanisms of Mutagenesis [4].
Table 1
Primer sequences for quantification of DNA methylation of DNMT3B and MTHFR genes using Sequenom EpiTYPER®. Upper case letters are used to denote the gene specific sequence in each primer shown.

| Gene | Fragment | Region | Primer sequences |
|------|----------|--------|------------------|
|      |          |        | **Forward 5’–3’** | **Reverse 5’–3’** |
| DNMT3b | 1 | chr20:31349428 to 31349886 | aggaagagagTTGAGTTATTTAGGATTTTGTT | cagtaatacactatagggagaagctCAGACTTCTCTAATACTTTTTC |
|       | 2 | chr20:31349921 to 31350352 | aggaagagagGAAAGGTTTTAGAGGCTTTT | cagtaatacactatagggagaagctTCTAACCCTAATGCTAAAAAC |
|       | 3 | chr20:31350385 to 31350876 | aggaagagagGTGGGTGTTGAAGGTTAGTTA | cagtaatacactatagggagaagctCCCCTCTCCATAAAGGAAAC |
|       | 4 | chr20:31350991 to 31351263 | aggaagagagGTTTTTGTTTTGGGAAAAGTTAAGG | cagtaatacactatagggagaagctCTAAAAATAAAAATAAAAAAC |
| MTHFR | 1 | chr1:11866175 to 11865734 | aggaagagagGAGTTTTGGGATTGAGATTAGGAGT | cagtaatacactatagggagaagctAAACAAAAAACCAATCTTC |
|       | 2 | chr1:11865705 to 11865291 | aggaagagagGAGATTTTTGTTTTTTT | cagtaatacactatagggagaagctCAACTAACCCTAATAACCTAA |
|       | 3 | chr1:11865198 to 11864850 | aggaagagagTAGGTTTTGTTTTTTTGTTT | cagtaatacactatagggagaagctAAATTTCTCATTCTTTCAAATCTTCA |
2. Experimental design, materials and methods

2.1. Overview of study design

Among the 330 patients, aged 40–65, undergoing a screening colonoscopy at Hotel Dieu Hospital in Kingston, Ontario, Canada, 272 participants were selected for this research as they met the minimum blood leukocyte DNA concentration required for DNA methylation analysis. Ninety participants were diagnosed with one or more pathologically-confirmed tubular, tubullo-villous or villous adenoma(s) while 182 patients had no abnormality identified during colonoscopy.

At the colonoscopy visit, a fasting blood sample was collected in a 10 ml ethylenediaminetetraacetic acid vacutainer. The vacutainer was placed immediately on ice and blood was centrifuged within 45 min; the buffy layer (blood leukocytes) was removed and stored at $-20^\circ$C. DNA was isolated from blood leukocytes and purified using the 5-PRIME DNA isolation kit (Inter Medico, Markham, ON, Canada). DNA concentrations were quantified using Quant-iT™ PicoGreen®, and DNA was stored at $-20^\circ$C until methylation analysis.

**Fig. 1.** Data cleaning and processing of methylation ratios for CpG sites within the *DNMT3B* and *MTHFR* genes.
2.2. Quantification of gene-specific DNA methylation

DNA methylation of DNMT3B and MTHFR genes was quantified in stored leukocytic DNA using the Sequenom EpiTYPER® technology. EpiTYPER® uses base-specific cleavage and matrix-assisted laser desorption/ionization-time of flight mass spectrometry to quantify DNA methylation [5–7]. This validated and reproducible method enables a highly precise and quantitative measurement of methylation at multiple CpG sites within large promoter regions of genes [6,7].

Methylation analysis was carried out at the McGill University and Génome Quebec Innovation Centre, Montreal, Quebec, Canada. Using Sequenom’s EpiDesigner™ software, PCR primers were designed for the two genes of interest (Table 1). Specifically, the genomic region of interest for DNMT3B spanned 1977 base pairs (GRCh37/hg19: chr20: 31349361-31351338, positive strand) and was chosen from the promoter region based on the reported location of a CpG island and literature showing variability between cancer cell lines [8–10]. The region of interest for MTHFR spanned 2115 base pairs (GRCh37/hg19: chr1:11866473-11864358, negative strand) and included coverage over the promoter region and CpG island; as well, CpG sites located apart from the island approximately 1000 base pairs from the promoter region were also targeted [10–14].

Participant DNA samples were bisulﬁte-treated and puriﬁed using the EZ DNA Methylation-Gold kit from ZymoResearch (cat D-5007) according to manufacturer’s protocol. This bisulﬁte conversion (BSC) step selectively deaminates unmethylated cytosines to uracil, but has no effect on 5-methylcytosines thereby leading to a primary DNA sequence change that permits differentiation of unmethylated cytosines from methylated cytosines.

![Fig. 2](image-url) (A) Average methylation ratio for each informative CpG site in DNMT3B according to chromosome location. CpG sites with Eigenvalues greater than 0.20 for PC1 and PC2 are indicated in red and green, respectively. (B) Diagram illustrating the transcription start site and exon 1 of the DNMT3b gene within the targeted CpG island. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
The regions of interest for DNMT3B and MTHFR were analyzed in 4 and 3 fragments spanning approximately 500 base pairs each, respectively. For each of the 7 fragments, 25 ng of BSC DNA was used to quantify methylation ratios within CpG units (a unit consists of either an individual CpG site or aggregates of multiple CpG sites) located within each fragment. A methylation ratio equals the percentage of methylated cytosines at a specific CpG unit of a gene, divided by the total number of copies of that CpG unit in the sample. For CpG units that consisted of multiple CpG sites, the methylation ratio of that CpG unit was assigned to each of the CpG sites within that unit. Three 96-well plates containing participant samples were run per fragment and for quality control, two high methylated human DNA controls and two low methylated human DNA controls manufactured by EpigenDx were included on each plate. Reliability was assessed using the two high methylated quality control samples and coefficients of variation (CV) were calculated between-plates; the low methylated DNA quality control sample was not utilized as the CV is sensitive to small changes to the mean when the mean is close to zero. For the high methylated DNA quality control sample, the CV between-plates was 5%.

2.3. Raw data processing and statistical analysis

Raw data consisted of methylation ratios for 147 CpG sites for DNMT3B and 83 CpG sites for MTHFR. Methylation data was cleaned separately for DNMT3B and MTHFR to remove methylation ratios for CpG sites that were unreliable and had little variability (SD ≤ 0.02) (Fig. 1). Specifically, unreliable methylation ratios for CpG sites were defined as either having: (1) low or high mass
outside of the limit of detection of the Sequenom Epityper assay or; (2) more than one silent peak or; (3) overlapping signals or; (4) duplicate CpG units of the same mass. Applying these criteria for DNMT3B, 50 of 147 methylation ratios for CpG sites were excluded. For MTHFR, 26 of 83 methylation ratios for CpG sites were excluded. CpG sites were subsequently defined as informative if they had methylation ratios with standard deviations > 0.02; this cut-off was applied to ensure that the calculated methylation measures were based on methylation ratios for CpG sites displaying meaningful differences. For DNMT3B, methylation ratios for 66 CpG sites had a standard deviation greater than 0.02. For MTHFR, methylation ratios for 28 CpG sites had a standard deviation greater than 0.02. These CpG sites were henceforth considered informative for statistical analysis. For DNMT3B, 31 of 66 informative CpG sites had complete data; for the remaining 35 CpG sites, methylation ratios for 1 or more CpG sites were missing for less than 11% of participants. Of the 28 informative CpG sites in MTHFR, data was complete for 4 CpG sites; for the remaining 24 CpG sites, methylation ratios for 1 or more CpG sites were missing for less than 4% of participants. All missing values were assigned the mean methylation ratio for each CpG site. To ensure that each participant had only a small proportion of missing values imputed, participants with data missing for > 10% of CpG sites within DNMT3B or MTHFR were excluded. Based on this cutoff, a total of 259 and 269 participants were included in the statistical analysis for DNMT3B and MTHFR, respectively.

For the statistical analysis, gene-specific DNA methylation was conceptualized using two approaches. First, measures representing average methylation across all informative CpG sites were created. Second, unsupervised principal component (PC) analysis was used as a pattern derivation and data-reduction approach, to develop two summary variables (PC1 and PC2) which together accounted for 32% and 42% of the variance of DNA methylation within informative CpG sites in DNMT3B and MTHFR, respectively.

2.4. Descriptive analysis of gene-specific DNA methylation in DNMT3B and MTHFR

The percent methylation of each informative CpG site for DNMT3B and MTHFR is presented in Figs. 2 and 3, respectively. The x-axis of the figures denotes where each informative CpG site resides with reference to chromosome location; the y-axis denotes the percent methylation of each informative CpG site. For DNMT3B, average percent methylation across all informative CpG sites among events and non-events was 11% (SD = 0.03) and 11% (SD = 0.04), respectively. For MTHFR, average percent methylation among events was 13% (SD = 0.03) while among non-events was 12% (SD = 0.02).

To evaluate possible patterns and correlations within the data, we differentiated CpG sites which correlated with each PC. Specifically, in Figs. 2 and 3, the CpG sites with Eigenvalues greater than 0.20 for PC1 are illustrated in red while those with Eigenvalues greater than 0.20 for PC2 are in green; CpG sites that did not have Eigenvalues greater than 0.20 for neither PCs are in gray. For both DNMT3B (Fig. 2) and MTHFR (Fig. 3), methylation ratios of CpG sites around the transcripational start site and the gene-coding area were correlated and comprised PC1 and PC2.

Acknowledgments

This work was supported by Colon Cancer Canada and support for Dr. V. Ho (postdoctoral fellow) was provided by the Canadian Institutes of Health Research Postdoctoral Fellowship.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2015.11.039.
References

[1] E. Taioli, M.A. Garza, Y.O. Ahn, D.T. Bishop, J. Bost, B. Budai, et al., Meta- and pooled analyses of the methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism and colorectal cancer: a HuGE-GSEC review, Am. J. Epidemiol. 170 (10) (2009) 1207–1221.

[2] S. Zhu, H. Zhang, Y. Tang, P. Liu, J. Wang, DNMT3B polymorphisms and cancer risk: a meta analysis of 24 case-control studies, Mol. Biol. Rep. 39 (4) (2012) 4429–4437.

[3] K.B. Michels, Introduction, in: K.B. Michels (Ed.), Epigenetic Epidemiology, Springer, New York, 2012, p. 1.

[4] V. Ho, J.E. Ashbury, S. Taylor, S. Vanner, W.D. King, Gene-specific DNA methylation of DNMT3B and MTHFR and colorectal adenoma risk, Mutat. Res. – Fundam. Mol. Mech. Mutagen. 782 (2015) 1–6.

[5] M. Ehrich, S. Zoll, S. Sur, D. van den Boom, A new method for accurate assessment of DNA quality after bisulfite treatment, Nucleic Acids Res. 35 (5) (2007) e29.

[6] M. Ehrich, M.R. Nelson, P. Stanssens, M. Zabeau, T. Liloglou, G. Xinarianos, et al., Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry, Proc. Natl. Acad. Sci. USA 102 (44) (2005) 15785–15790.

[7] M.W. Coolen, A.L. Statham, M. Gardiner-Garden, S.J. Clark, Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements, Nucleic Acids Res. 35 (18) (2007) e119.

[8] C. Huidobro, R.G. Urdinguio, R.M. Rodriguez, C. Mangas, V. Calvanese, P. Martinez-Camblor, et al., A DNA methylation signature associated with aberrant promoter DNA hypermethylation of DNMT3B in human colorectal cancer, Eur. J. Cancer Oxf. Eng. 1990 48 (14) (2012) 2270–2281.

[9] G. Rajendran, K. Shanmuganandam, A. Bendre, D. Muzumdar, A. Goel, A. Shiras, Epigenetic regulation of DNA methyltransferases: DNMT1 and DNMT3B in gliomas, J. Neurooncol. 104 (2) (2011) 483–494.

[10] L.R. Meyer, A.S. Zweig, A.S. Hinrichs, D. Karolchik, R.M. Kuhn, M. Wong, et al., The UCSC Genome Browser database: extensions and updates 2013, Nucleic Acids Res. 41 (2013).

[11] ENCODE Project Consortium E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigo, T.R. Gingeras, et al., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature 447 (7146) (2007) 799–816.

[12] S.E. Celniker, L.A. Dillon, M.B. Gerstein, K.C. Gunsalus, S. Henikoff, G.H. Karpen, et al., Unlocking the secrets of the genome, Nature 459 (7249) (2009) 927–930.

[13] T. Vaissiere, R.J. Hung, D. Zaridze, A. Moukeria, C. Cuenin, V. Fasolo, et al., Quantitative analysis of DNA methylation profiles in lung cancer identifies aberrant DNA methylation of specific genes and its association with gender and cancer risk factors, Cancer Res. 69 (1) (2009) 243–252.

[14] S. Mani, K. Szymanska, C. Cuenin, D. Zaridze, K. Balassiano, S.C. Lima, et al., DNA methylation changes associated with risk factors in tumors of the upper aerodigestive tract, Epigenet. J. DNA Methylation Soc. 7 (3) (2012) 270–277.