Journal of Computational Biology and Bioinformatics Research Vol. 2(2), pp. 20-22, October 2010
Available online http://www.academicjournals.org/jcbbr
ISSN-2141-2227 ©2010 Academic Journals

Short Communication

SNP model to address cytosine trios

Sriram Kannan

Graduate Studies, Molecular Mechanism of Disease, NCMLS, Radboud University, K603, Erasmuslaan17, 6525GE, Nijmegen, Netherlands. E-mail: sriram.kannan@student.ru.nl. Tel: 0031-649775011.

Accepted 27 August, 2010

DNA methylation maintains allele specific gene expression (Chan et al., 2003) in which miRNA influences allele-specific protein expression and SNPs, found inside miRNA, in turn influences tumor susceptibility (Nicoloso et al., 2010). Further, methylated CpGs have been correlated to *APC* gene in colorectal cancer (Zhang et al., 2007) and retrotranspositions have also been correlated to *APC* gene (Miki et al., 1992). Another aspect unrelated to it is *Tet1* gene, which has been associated with the conversion of methylcytosine to hydroxymethyl cytosine (Tahiliani et al., 2009). DNA methylation might also have a role in the prevention of normal differentiation in pediatric cancers (Diede et al., 2009). As the difference between a nucleobase and its methylated form is its structure and its molecular weight, this research article is focussed on using the molar mass of nucleobases to find out if there is any uniqueness as for the position of occurrence of a nucleotide in a given model. SNPs occurrence position was used as a model in this research for addressing the cytosine trios (cytosine, 5methyl cytosine, hydroxy methyl cytosine) based on molar mass of the nucleobase. As the conditions for occurrence of SNP, at a given position in a sequence, were found to uniformly conform with all 140,000 SNPs analysed, including all clinically associated SNPs from NCBI SNP database, it intrigued conformity to be cross checked with SNPs near experimentally proven methylation sites.

Key words: SNP, molar mass, *Tet1,* colorectal cancer, methylation.

INTRODUCTION

SNPs could influence miRNA as seen with rs334348 associated with germline allele specific expression of *TGFBR1* correlated to colorectal cancer (Nicoloso et al., 2010). Methylated cytosines have been experimentally proven in CpG islands of *APC* gene promoter region (Genbank accession: U02509, CpG Position 687) (Zhang et al., 2007) and rs35417795, which is another SNP, occurs at close proximity in this island. Further, the tumor suppressor gene, *Tet1*, is known for its conversion of methylated cytosine to hydroxymethylcytosine.

Hence, the intriguing question was to find out if there are SNP occurrence positions that might be influenced by methylation positions. The first step was to find out if there are any global SNP occurrence positions, while the second step was to find out if this could be changed by substituting methylated cytosine molar mass instead of cytosine and subsequently, with hydroxy methyl cytosine as *Tet1* does.

After analyzing, all the validated biallelic SNPs of chromosome 21, X, Y, mitochondrial and all the clinically associated SNPs of the other chromosomes by the novel, excel the based algorithm as described in the methods that conform globally to the occurrence position of SNPs (Table 2). This pattern was used to find out if any deviation occurs, when the experimentally proven methylated cytosine was substituted in flanking regions of SNP occurrence and then replaced by hydroxymethylcytosine.

MATERIALS AND METHODS

Global SNP occurrence pattern

The algorithm takes an input of a DNA sequence as:

"……N_a N_b N_c N_d N_e N_f N_g N_h N_i N_j N_k N_l N N_m N_n N_o N_p N_q N_r N_s N_t N_u N_v N_w ....."

where N is the position of interrogation of being x or y nucleotide representing a single nucleotide change and {a, b, c, d, e, f, g, h, I, j, k, X/Y, m, n, o, p, q, r, s, t, u, v and w) are the positions of nucleobases in an ascending order. It then substitutes the nucleobases nucleobase occurrence position with a numerical constant value "NC" for each base as shown in Table 1 substitution values. After the substitution with any one type of NC, there are four ratio values that are calculated (R1, R2, R3 and R4) as shown in Table 2. The ratio value is calculated with this formula:

Table 1. Molar mass based on substitution values.

	A	G	C	T	Methyl cytosine	Hydroxy methyl cytosine
Type one substitution	24	40	1	15	13	30
Logic used	Difference in molar mass from cytosine with decimals rounded off with an addition of one so as not to assign cytosine a zero value					
Name assigned	Nucleobase value (NBV)					
Type two substitution	23.03	39.03	1	13.94	13.19	30.03
Logic used	Absolute difference in molar mass from cytosine					
Name assigned	Absolute NBV					
Type three substitution	135.13	151.03	111.1	126.04	125.29	141.13
Logic used	Absolute molar mass value (1)					
Name assigned	Absolute value					

Table 2. Global SNP occurrence positions with type 1 substitutions.

Set	R1(NC1.5)/(NC1.5)	R2(NC1.5)/(NC1.5)	R3(NC1.5)/(NC1.5)	R4(NC1.5)/(NC1.5)
Set 1	a,b,c,d,e / g,h,i,j,k	b,c,d,e,f / h,i,j,k,x	a,b,c,d,e / g,h,i,j,k	b,c,d,e,f / h,i,j,k,y
Set 2	b,c,d,e,f / h,i,j,k,x	c,d,e,f,g / i,j,k,x,m	b,c,d,e,f / h,i,j,k,y	c,d,e,f,g / i,j,k,y,m
Set 3	c,d,e,f,g / i,j,k,x,m	d,e,f,g,h / j,k,x,m,n	c,d,e,f,g / i,j,k,y,m	d,e,f,g,h / j,k,y,m,n
Set 4	d,e,f,g,h / j,k,x,m,n	e,f,g,h,i / k,x,m,n,o	d,e,f,g,h / j,k,y,m,n	e,f,g,h,i / k,y,m,n,o
Set 5	e,f,g,h,i / k,x,m,n,o	f,g,h,i,j/ x,m,n,o,p	e,f,g,h,i / k,y,m,n,o	f,g,h,i,j/ y,m,n,o,p
Set 6	f,g,h,i,j/ x,m,n,o,p	g,h,i,j,k/ m,n,o,p,q	f,g,h,i,j/ y,m,n,o,p	g,h,i,j,k/ m,n,o,p,q
Set 7	g,h,i,j,k/ m,n,o,p,q	h,i,j,k,x / n,o,p,q,r	g,h,i,j,k/ m,n,o,p,q	h,i,j,k,y / n,o,p,q,r
Set 8	h,i,j,k,x / n,o,p,q,r	i,j,k,x,m/ o,p,q,r,s	h,i,j,k,y / n,o,p,q,r	i,j,k,y,m/ o,p,q,r,s
Set 9	i,j,k,x,m/ o,p,q,r,s	j,k,x,m,n / p,q,r,s,t	i,j,k,y,m/ o,p,q,r,s	j,k,y,m,n / p,q,r,s,t
Set 10	j,k,x,m,n / p,q,r,s,t	k,x,m,n,o /q,r,s,t,u	j,k,y,m,n / p,q,r,s,t	k,y,m,n,o /q,r,s,t,u
Set 11	k,x,m,n,o /q,r,s,t,u	x,m,n,o,p /r,s,t,u,v	k,y,m,n,o /q,r,s,t,u	y,m,n,o,p /r,s,t,u,v
Set 12	x,m,n,o,p /r,s,t,u,v	m,n,o,p,q /s,t,u,v,w	y,m,n,o,p /r,s,t,u,v	m,n,o,p,q /s,t,u,v,w

Set	Condition (Any one of these)	SNP	Conditions
1,6,7,12	R1 = R2 = R3; R1 = R3 = R4; R2 = R3 = R4; R2 = R4 = R1	+	1
2,3,4,5,8,9,10,11	R1 = R2 and R3 = R4; R1 = R2 and R3≠R4; R3 = R4 and R1≠R2	+	2
1,6,7,12	R1 = R2 and R3 = R4; R1 = R2 and R3 ≠ R4; R3 = R4 and R1 ≠ R2	-	3
2,3,4,5,8,9,10,11	R1 = R2 = R3; R1 = R3 = R4; R2 = R3 = R4; R2 = R4 = R1	-	4
All sets	R1 = R2 = R3 = R4	-	5
All sets	All four Rs are not equal	+	6

$R = \sum (NC1.5) / \sum (NC1.5)$,

Where NC1.5 depicts: (NC1 + NC2 + NC3 + NC4 + NC5) as for positions mentioned in Table 2.

The positions are dynamic as the algorithm calculates “NC” along the sequence for positions in the sequence {a, b, c, d, e, f, g, h, I, j, k, X/Y, m, n, o, p, q, r, s, t, u, v and w) including only five positions for one set as in Table 1, whereas there are twelve sets calculated for each R value altogether. Based on the four R values for the twelve sets, six different conditions were concluded as for the occurrence positions of SNPs that are summarised in Table 2. R1 and R2 calculates for positions with “x” as nucleobase at position “N”, while R3 and R4 calculates for positions with “y” as nucleobase at position “N”. The DNA substitutions were simulated with the following values as in Table 1. Table 2 values are based on type one substitutions.

RESULTS AND DISCUSSION

The global SNP occurrence position as in Table 2 was found to conform with 140,000 SNPs analysed, but not with rs35417795 of *APC* gene promoter region lying close to the methylated CpG island if anlaysed with molar mass of methylated cytosine instead of cytosine.

This could mean that protein encoded genes like *tet1* could probably use this strategy of hydroxymethyl cytosine conversion from methylated cytosine to prevent the occurrence of a random nucleotide change in a nearby flanking region which could be pathogenic sometimes. This model could serve as a tool for cancer researchers to validate this experimentally.

Key points

(1) SNP occurrence positions are correlated with methyl cytosine and a tumor suppressor gene function.
(2) In special attention to colorectal cancer related attributes.

REFERENCES

Chan HW, Kurago ZB, Stewart CA, Wilson MJ, Martin MP, Mace BE, Carrington M, Trowsdale J, Lutz CT (2003). DNA methylation maintains allele-specific KIR gene expression in human natural killer cells. J. Exp. Med., Jan 20; 197(2):245-55.

Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res., Feb 1; 52(3):643-5.

Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, Shimizu M, Wojcik SE,Ferdin J, Kunej T, Xiao L, Manoukian S, Secreto G, Ravagnani F, Wang X, Radice P,Croce CM, Davuluri RV, Calin GA (2010). Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. Cancer Res., Apr 1; 70(7):2789-98.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S,Iyer LM, Liu DR, Aravind L, Rao A (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science, May 15; 324(5929):930-5.

Zhang LX, Pan SY, Chen D, Xie EF, Gao L, Shu YQ, Lu ZH, Cheng L,Yang D,Zhang JN (2007). Effect of adenomatous polyposis coli (APC) promoter methylation on gene transcription in lung cancer cell lines. China J.Cancer, 26(6):576-80.