Comparison of Methylation Density in Different Cancer Types by Illumina Infinium HumanMethylation450 Methods

Şenol Doğan[1], Hakan Şahin[2]

1-Department of Genetics and Bioengineering, International Burch University
Francuske revolucije bb, Ilidža 71000, Sarajevo, Bosnia and Herzegovina
2-Department of Information Technology, International Burch University
Francuske revolucije bb, Ilidža, 71000, Sarajevo, Bosnia and Herzegovina

Abstract.

Abnormal hypermethylation of DNA is highly detected in many cancer types. It is accepted that the hypermethylation contributes to tumor developing by silencing the tumor suppressor genes expression. New data techniques, Illumina Infinium HumanMethylation450, gives us a wide approach about genome methylation density and location. 8 different TCGA cancer data were compared. As a result, hypomethylated and hypermethylated genes were selected. Moreover, the highest methylated chromosomes, such as 19, 16, 17, 7, 11, 1, have been detected among the cancer types. The lowest and the highest beta value of the genes and chromosomes may give us a methylation range to understand future methylation process. Physical map of hypermethylated, near the both ends of chromosomes, and hypomethylated genes, near the centromere, is another result of this paper.

***Keywords**:* Hyper and Hypomethylation, TCGA, Illumina Infinium HumanMethylation450, Methylation location of genes and chromosomes.

1. Introduction

DNA is not a stable molecule. It is active, productive and developmentally changing one. Replication and productivity makes it unique molecule. Epigenetic is one of the main reason of DNA variation. Aging, diseases, organ development and changing organism by the time are regulated by epigenetic regulation mechanism [1][2]. The main source of epigenetics is methylation. Addition of methyl (-CH3) group to cytosine bases located in (CpG) islands, is the source of DNA development [3]. Cancer types [4], ageing and stem cell profiling, silencing of repetitive elements, inactivation of X chromosomes, RNA interference, imprinting, specific tissue expression levels are described the impact of specific DNA methylation patterns [5][10]. The latest results show that DNA methylation and histone modification have very close functional relationship [11]. DNA hypermethylation occurs especially among CpG island of promoters in many cancer types [12]. Hypermethylated promoter sites causes silencing the affected genes expression [13]. It is widely used to

identify DNA methylation of oncogenes, upstream, enhancer, promoters, in various cancer types [14].There is a general acceptance that abnormal methylation of promoter sites is an important cause in alteration of gene expression on cancer [15]. Using high-throughput profiling of CpG islands methylation is the essential way for further understanding of the influence of epigenomics [16]. Illumina Infinium 450 Methylation techniques give us the newest and more detailed data to understand potential cancer gene methylation density and location. Even though there are some statistical calculations Illumina recommends to use a Beta-value, ranging from 0 to 1, measuring the percentage of methylation [17]. TCGA database provides us the latest data for hundreds of discrete cancer type's methylation profiles [18]. Discovery of methylation biomarkers could be affected by batch effect [14], but it is neglected in this paper.

2. Materials and Methods

2.1 Bisulfite sequencing

Infinium HumanMethylation27 (27k) BeadChip technology was early used technology to measure the human methylation. Recently, the genomic coverage of the array has increased tremendously. So it gives a production of the Infinium HumanMethylation450BeadChip, which contains the methylation examples of 485,577 CpG in the human genome. The bisulfite conversion of unmethylated cytosine's technique generates Infinium methylation probes query a C/T polymorphism in the genome. A combination of two distinct probe types, Infinium I and II , is the newest way to be used by the Infinium 450k methylation platforms [19]. Bisulfite sequencing is the use of bisulfite treatment of DNA to determine its pattern of methylation. DNA methylation was first discovered in epigenetic mark. Quantitative and qualitative analysis of CpG methylation are being well studied over the last years. The bisulfite treatment is the conversion of non-methylated cytosine to uracil and methylated cytosine to cytosine. Analyzing this kind of data is complicated. Alignment of bisulfite treated sequence to reference sequence are necessary before getting methylation profiles for future statistical analysis [20].

2.2 Data Sources

The batches of HumanMethylation450 have different data levels, but Level 3 data sets were used for this paper **(Table 1)**. Data sets from TCGA were categorized and compared as T, TN, NT, and N **(Table 2)**. Gene Symbol, Chromosome Number, Location and Beta-value were analyzed to distinguish the highly methylated genes and location.

Table 1. Meaning of TCGA data levels

Batch Level	Type
Level 1	Raw signal intensities of probes
Level 2	Calculated beta values
Level 3	Calculated beta values mapped to genome, per sample

Table 2. Abbreviation of TCGA data

Abbreviation	T	TN	N	NT
Meaning	Tumor No Matched	Tumor Matched	Normal, No Matched	Normal Matched

2.3 Calculation of Methylation Level

Beta Value and M-value are the methods of quantifying methylation levels. First method is Beta value, from 0 to 1, generally used to quantify the methylation percentage. M-value is another method and log2 ratio is the way, methylated towards unmethylated probes intensities [2].

Beta-value

$$\boldsymbol{Beta_i} = \frac{\mathbf{max}(y_{i,methy},\mathbf{0})}{\mathbf{max}(y_{i,unmethy},\mathbf{0})+\mathbf{max}(y_{i,methy},\mathbf{0})+\boldsymbol{\alpha}} \quad (1)$$

M-value

$$M_i = log_2\left(\frac{\max(y_{i,methy},0)+\alpha}{\max(y_{i,unmethy},0)+\alpha}\right) \quad (2)$$

For the calculation we might use Beta value (1) and M-value (2) together but Illumina recommends Beta–value. The Beta-value of the probes are calculated with the formula of *Beta* = M / (U + M) +100 where *Beta* $\in$ [0, 1] M and U stand for Methylated and Unmethylated respectively.

3. Experiments

Eight different cancer types, BRCA, COAD, GBM, KICH, LIHC, READ, PAAD, STAD, were chosen from TCGA database (Table 3). Depending on their categories T, TN, NT, N, 298 different available samples (Table 3) are analyzed for this paper. The chosen cancer types methylation levels searched and analyzed. The second step was

to compare the data according to their category beta value. Finally genes and chromosomes are tabled depending on their highest and lowest beta value. The beta value of the genes that higher than 0.99 were chosen for the experiments.

4. Results

4.1 The Methylation Range

As a result of the whole experiments (Table 3) the cancer methylation range was generated. The highest methylated gene ARID3A, beta value 0.996, and the lowest methylated gene KLF10, beta value 0.0032, are found in READ and BRCA respectively (Figure1). The two values show the methylation border from lowest to the highest.

Table 3. Experimental Samples for each Cancer type

Cancer type	Abbreviation	T	TN	N	NT	Total Samples
Breast Invasive Carcinoma	BRCA	22	13	8	17	60
Colon adenocarcinoma	COAD	25	12	N/A	16	53
Glioblastoma Multiforme	GBM	26	1	N/A	1	28
Kidney renal clear cell carcinoma	KIRC	14	16	1	20	51
Liver hepatocellular carcinoma	LIHC	4	18	N/A	2	24
Rectum Adenocarcinoma	READ	12	6	N/A	N/A	18
Pancreatic adenocarcinoma	PAAD	24	7	N/A	7	38
Stomach adenocarcinoma	STAD	24	2	N/A	N/A	26
Total Samples	8	151	75	9	63	298

4.2.1 Hypermethylated Genes

Each cancer type has different number of methylated genes. Some of the chromosomes carry high methylated genes (Table 4). Ten common highest methylated genes are detected, but first five are presented here. According to the highest methylation level, FBRSL1, C1QTNF8, ANKRD11, DAZAP1 and CASP9 genes are commonly found for the cancer types. The Beta values of the genes are all bigger than 0.99. ARID3A, beta value 0.996, is the highest methylated gene among all. This is a tremendous methylation value for a gene.

Table 4. Top five hypermethylated genes and their beta value according to cancer types

Cancer Types	12;FBRSL1	19;DAZAP1	16;C1QTNF8	1; ASP9	16; NKRD11
BRCA	0.9957	0.9954	0.9952	0.9950	0.9943
COAD	0.9940	0.9933	0.9931	0.9931	0.9937
GBM	0.9953	0.9948	0.9942	0.9941	0.9940
KICH	0.9940	0.9936	0.9782	0.9909	0.9944
LIHC	0.9955	0.9951	0.9942	0.9943	0.9941
PAAD	0.9939	0.9942	0.9929	0.9925	0.9926
READ	0.9923	0.9944	0.9908	0.9942	0.9938
STAD	0.9939	0.9938	0.9936	0.9936	0.9939

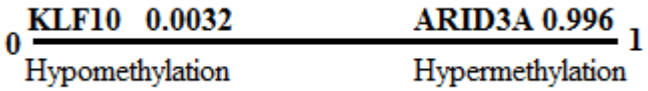

Fig. 1. The methylation range among the cancer types

4.2.2 Hypermethylated Genes Location

Hypermethylated genes locations on the chromosomes have been searched [21] and they show similar characteristic. All hypermethylated genes locate the both ends of chromosomes (Fig. 2). It clearly shows that ends of chromosomes are under high risk of methylation in the cancer types.

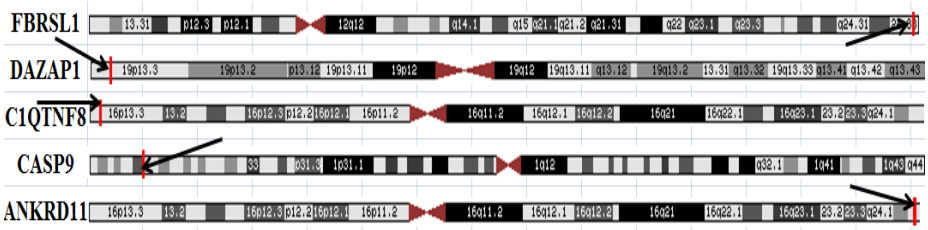

Fig. 2. Five common hypermethylated genes location on chromosomes. The arrows for each chromosomes show position of the genes. The hypermethylated genes locate near p or q ends.

4.3.1 Hypomethylated Genes

There are also five common lowest methylated genes found. The lowest beta value belongs to KLF10 gene with 0.003 from BRCA. RPL19, BCL2L11, ICT1, IFT88 and TTC39C (Table 5) genes are commonly hypomethylated genes for the cancer types.

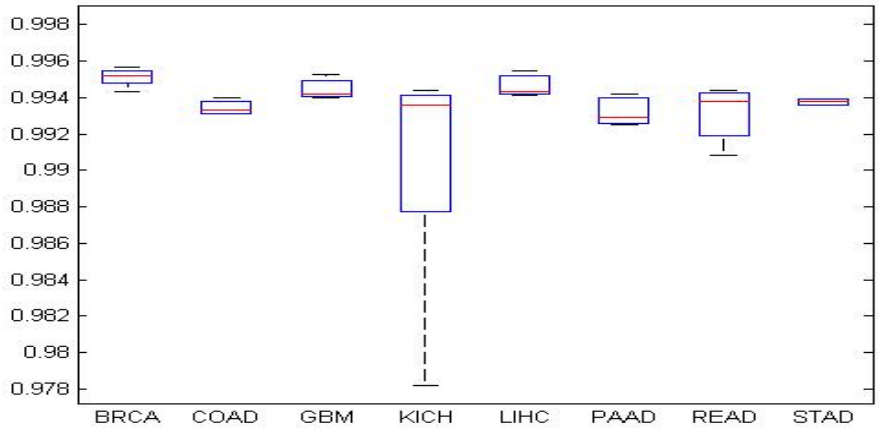

Fig. 3. Top five hypermethylated genes, FBRSL1, DAZAP1, C1QTNF8, CASP9, ANKRD11 comparison range.

Table 5 . Top lowest hypomethylated genes and their beta value according to cancer types

Cancer Types	17; RPL19	2;BCL2L111	17; ICT1	13;IFT88	18;TTC39C
BRCA	0.0047	0.0050	0.0051	0.0054	0.0054
COAD	0.0058	0.0060	0.0073	0.0070	0.0064
GBM	0.0046	0.0056	0.0053	0.0052	0.0054
KICH	0.0059	0.0070	0.0091	0.0076	0.0064
LIHC	0.0053	0.0057	0.0059	0.0062	0.0054
PAAD	0.0064	0.0058	0.0060	0.0071	0.0067
READ	0.0045	0.0050	0.0054	0.0045	0.0054
STAD	0.0058	0.0060	0.0068	0.0072	0.0061

4.3.2 Hypomethylated Genes Location

The five common genes were searched by the same genome browser tool [21]. Oppositely hypermethylated genes location, hypomethylated genes locate closer to centromere (Fig. 4). One of the ten lowest methylated gene is away from the centromere ICT1. The rest are all very closer position.

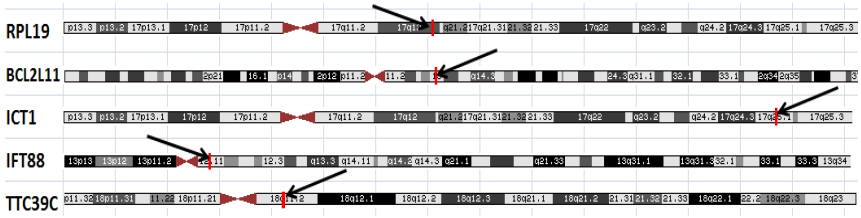

Fig. 4. The five Hypomethylated genes location on chromosomes. The arrows show the position of the hypomethylation genes. The hypomethylated genes locate near the centromere.

4.4 Higher Methylated Gene Number in the Chromosomes

19,16,7,11,1,17,8 and 12 chromosomes` hypermethylation levels are apparently bigger than other chromosomes. The chromosomes contain of high amount of hypermethylated genes, beta value higher than 0.99 (Table 6). Interestingly, the eight chromosomes are the common top hypermethylated for the 8 different cancer. It is understood that some chromosomes carry so sensitive genes for methylation.

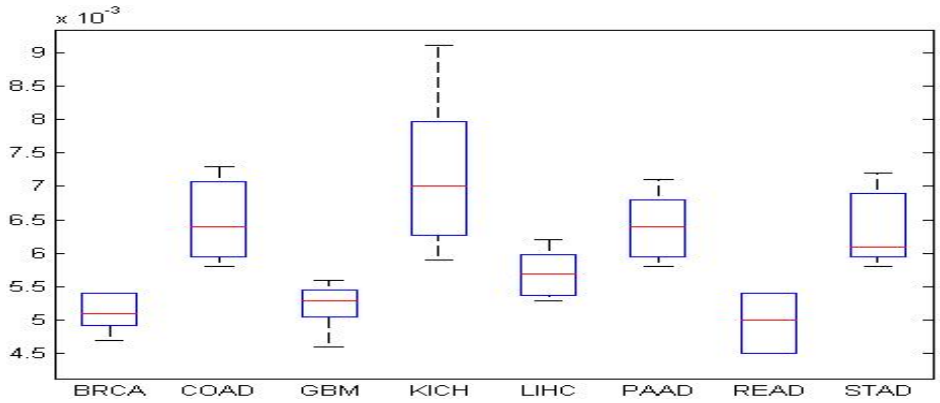

Fig. 5. Top five hypomethylated genes, RPL19, BCL2L111, ICT1, IFT88, TTC39C comparison range.

Table 6. Chromosome No & Respective methylated gene # per sample

BRCA	COAD	GBM	LIHC	KICH	PAAD	READ	STAD
16 ; 76.9	19 ; 40.8	19 ; 54.1	16 ; 60.9	16 ; 81.1	19 ; 80.7	16 ; 131.4	19 ; 27
19 ; 73.6	07 ; 30.9	07 ; 48	19 ; 57.7	19 ; 75.3	16 ; 80.6	19 ; 116.4	16 ; 24.8
07 ; 58.5	16 ; 26.9	16 ; 42.5	07 ; 48.5	07 ; 61.2	07 ; 65.5	07 ; 98.8	17 ; 20.6
01 ; 56	17 ; 26.6	17 ; 35.5	17 ; 42.7	11 ; 47.7	17 ; 58	17 ; 79.4	07 ; 18
17 ; 54.5	01 ; 20.4	01 ; 33.7	01 ; 39.9	08 ; 31.5	01 ; 54.5	01 ; 73	01 ; 14.8
11 ; 41.1	11 ; 17.6	11 ; 26.6	11 ; 34	17 ; 29.7	11 ; 44	11 ; 64	08 ; 13.2
08 ; 35.6	08 ; 15	08 ; 19.9	08 ; 26.9	12 ; 29.6	08 ;36.4	08 ; 56.6	11 ; 10
12 ; 29.8	12 ; 14.7	12 ; 18.1	12 ; 24.6	01 ; 29.4	02 ; 34.2	12 ; 48.2	12 ; 9

5. Conclusions

To build a cancer specific biomarker is very hard for each cancer type [14]. Methylation map can give us a new approach to analyze the cancer pathway. The methylation range is between KLF10, 0.0032, and ARID3A, 0.996, among selected cancer types. There are some genes and chromosomes highly dominated to be hypermethylated or hypomethylated (Table 4 and 5). Especially, 8 chromosomes of the cancer types are same about hypermethylation. Table 6 shows the hypermethylated genes number of the top eight chromosomes. So the hypermethylation of chromosomes draw us to new idea to understand epigenetically

changing of DNA and cancer alteration. The main goal of the paper is to discover the high and low methylated genes and chromosomes in cancer types which depending on the latest next-generating sequences technologies.

References

1. Laird PW: Principles and challenges of genome-wide DNA methylation analysis. Nat Rev Genet 2010, 11(3):191-203.
2. Pan Du, Xiao Zhang, Chiang-Ching Huang, Nadereh Jafari, Warren A Kibbe,Lifang Hou and Simon M Lin: Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis, BMC Bioinformatics 2010, 11:587 doi:10.1186/1471-2105-11-587
3. Lutz P. Breitling, Rongxi Yang, Bernhard Korn, Barbara Burwinkel, and Hermann Brenner; Tobacco-Smoking-Related Differential DNA Methylation: 27 K Discoveries and Replication, Am J Hum Genet. 2011 April 8; 88(4): 450–457.doi: 10.1016/j.ajhg.2011.03.003
4. Feinberg AP, Tycko B: The history of cancer epigenetics. Nat Rev Cancer 2004, 4:143-153.
5. Reik W, Santos F, Dean W: Mammalian epigenomics: reprogramming the genome for development and therapy. Theriogenology 2003, 59:21.32.
6. Yoder JA, Walsh CP, Bestor TH: Cytosine methylation and the ecology of intragenomic parasites.Trends Genet 1997, 13:335-340.
7. Chow JC, Yen Z, Ziesche SM, Brown CJ: Silencing of the mammalian X chromosome. Annu Rev Genomics Hum Genet 2005, 6:69-92
8. Hawkins PG, Morris KV: RNA and trans. Mod. of gene exp.. Cell Cycle 2008, 7:602-607.
9. Morison IM, Ramsay JP, Spencer HG: A census of mammalian imprinting. Trends Genet 2005, 21:457-465.
10. Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, Tomazou EM, Backdahl L, Johnson N, Herberth M, et al.: An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). Genome Res 2008, 18:1518-1529.
11. Cedar H, Bergman Y: Linking DNA methylation and histone modification: patterns and paradigms.Nat Rev Genet 2009, 10:295-304.
12. Duncan Sproul , Robert R Kitchen , Colm E Nestor, J Michael Dixon, Andrew H Sims, David J Harrison, Bernard H Ramsahoye and Richard R Meehan; Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. Genome Biology 2012, 13:R84 doi:10.1186/gb-2012-13-10-r84
13. Jones PA, Baylin SB: The fundamental role of epigenetic events in cancer. Nat Rev Genet 2002, 3:415-428.
14. Yao C, Li H, Shen X, He Z, He L, et al. (2012) Reproducibility and Concordance of Differential DNA Methylation and Gene Expression in Cancer. PLoS ONE 7(1): e29686. doi:10.1371/journal.pone.0029686
15. Kanduri M, Cahill N, Goransson H, Enstrom C, Ryan F, et al. (2010) Differential genome-wide array-based methylation profiles in prognostic subsets of chronic lymphocytic leukemia. Blood 115: 296–305.
16. Bell CG, Teschendorff AE, Rakyan VK, Maxwell AP, Beck S, Savage DA: Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. BMC Med Genomics 2010, 3:33.
17. .Bibikova M, Fan JB: GoldenGate assay for DNA methylation profiling.Methods Mol Biol 2009, 507:149-163.
18. (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455: 1061–1068.
19. Maksimovic, J; Gordon, L. ;Oshlack, A. (2012) SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips, *Genome Biology* 2012, **13**:R44 doi:10.1186/gb-2012-13-6-r44
20. Arie Zackay1,2, Christine Steinhoff1*Meth Visual - visualization and exploratory statistical analysis of DNA methylation profiles from bisulfite sequencing
21. http://genome.ucsc.edu/